A.A. SAMARSKII

MÉTODOS DE SOLUCÍON DE LAS ECUACIONES RETICULARES

1

A.A. Samarski
E.S. Nikolaev

Métodos de solución de las ecuaciones reticulares

Tomo

I

Editorial Mir
Moscú

Contenido

Prólogo 9

Introducción 13

Capítulo I. Métodos directos de solución de ecuaciones en diferencias 30

§ 1. Ecuaciones reticulares. Conceptos fundamentales . . . 30

1. Retículos y funciones reticulares (30). 2. Derivadas de diferencias y algunas identidades de diferencias (33). 3. Ecuaciones reticulares y en diferencias (38). 4. Problema de Cauchy y problemas de contorno para las ecuaciones en diferencias (42).

§ 2. Teoría general de las ecuaciones lineales en diferencias 46

1. Propiedades de las soluciones de una ecuación homogénea (46). 2. Teoremas sobre las soluciones de una ecuación lineal (50). 3. Método de variación de las constantes (52). 4. Ejemplos (57).

§ 3. Solución de ecuaciones lineales con coeficientes constantes 61

1. Ecuación característica. Caso de raíces simples (61). 2. Caso de raíces múltiples (63). 3. Ejemplos (66).

§ 4. Ecuaciones de segundo orden con coeficientes constantes 70

1. Solución general de una ecuación homogénea (70). 2. Polinomios de Chebishev (72). 3. Solución general de una ecuación no homogénea (75).

§ 5. Problemas de diferencias sobre valores propios 80

1. Primer problema de contorno en valores propios (80). 2. Segundo problema de contorno (83). 3. Problema de contorno mixto (84). 4. Problema de contorno periódico (87).

Capítulo II. Método de factorización 92

§ 1. Método de factorización para ecuaciones tripuntuales 92

1. Algoritmo del método (92). 2. Método de las factorizaciones opuestas (96). 3. Fundamentación del método de factorización (98). 4. Ejemplos de aplicación del método de factorización (102).

§ 2. Variantes del método de factorización 106

1. Variante por flujos del método de factorización (106). 2. Método de factorización cíclica (109). 3. Método de factorización para sistemas complejos (114). 4. Método de factorización no monótona (118).

§ 3. Método de factorización para ecuaciones pentapuntuales 123

1. Algoritmo de factorización monótona (123). 2. Fundamentación del método (126). 3. Variante de factorización no monótona (128).

§ 4. Método de factorización matricial 130

1. Sistemas de ecuaciones vectoriales (130). 2. Factorización para ecuaciones vectoriales tripuntuales (135). 3. Factorización para ecuaciones vectoriales bipuntuales (139). 4. Factorización ortogonal para ecuaciones vectoriales bipuntuales (142). 5. Factorización para ecuaciones tripuntuales con coeficientes constantes (147).

Capítulo III. Método de reducción total 153

§ 1. Problemas de contorno para ecuaciones vectoriales tripuntuales 153

1. Planteamiento de los problemas de contorno (153). 2. Primer problema de contorno (155). 3. Otros problemas de contorno para ecuaciones en diferencias (159). 4. Problema de Dirichlet de diferencias con orden de exactitud aumentado (162).

§ 2. Método de reducción completa para el primer problema de contorno 165

1. Proceso de exclusión impar-par (165). 2. Transformación del segundo miembro e inversión de las matrices (168).

3. Algoritmo del método (172). 4. Segundo algoritmo del método (172).

§ 3. Ejemplos de aplicación del método 182

1. Problema de Dirichlet de diferencias para la ecuación de Poisson en un rectángulo (182). 2. Problema de Dirichlet de diferencias con orden de exactitud aumentado (186)

§ 4. Método de reducción total para otros problemas de contorno 189

1. Segundo problema de contorno (189). 2. Problema periódico (196). 3. Tercer problema de contorno (200)

Capítulo IV. Método de separación de variables 209

§ 1. Algoritmo de la transformación de Fourier discreta . . 209

1. Planteamiento del problema (209). 2. Desarrollo en senos y en senos desplazados (214). 3. Desarrollo en cosenos (224). 4. Transformación de una función real periódica reticular (226). 5. Transformación de una función compleja periódica reticular (232).

§ 2. Resolución de problemas de diferencias por el método de Fourier 235

1. Problemas de diferencias sobre valores propios para el operador de Laplace en un rectángulo (235). 2. Ecuación de Poisson en un rectángulo. Desarrollo en serie doble (241). 3. Desarrollo en serie de aspecto singular (246)

§ 3. Método de reducción incompleta 251

1. Combinación de los métodos de Fourier y de reducción (251). 2. Resolución de los problemas de contorno para la ecuación de Poisson en un rectángulo (259). 3. Problema de Dirichlet de diferencias de alto orden de exactitud en un rectángulo (263).

Capítulo V. Aparato matemático de la teoría de los metodos iterativos 269

§ 1. Algunos conocimientos del análisis funcional 269

1. Espacios lineales (269). 2. Operadores en espacios lineales normados (273). 3. Operadores en un espacio de Hilbert (276). 4. Funciones de un operador acotado (283). 5. Operadores en un espacio de dimensión finita (284). 6. Solubilidad de las ecuaciones operacionales (287)

§ 2. Esquemas de diferencias como ecuaciones operacionales 290

1. Ejemplos de espacios de funciones reticulares (290). 2. Algunas identidades de diferencias (294). 3. Cotas de los operadores de diferencias más simples (297). 4. Estimaciones inferiores para algunos operadores de diferencias (301). 5. Estimaciones superiores para operadores de diferencias (309). 6. Esquemas de diferencias como ecuaciones operacionales en espacios abstractos (311). 7. Esquemas de diferencias para ecuaciones elípticas de coeficientes constantes (315). 8. Ecuaciones de coeficientes variables y con derivadas mixtas (318)

§ 3. Conceptos fundamentales de la teoría de los métodos iterativos 324

1. Método de establecimiento (324). 2. Esquemas iterativos (325). 3. Convergencia y número de iteraciones (328). 4. Clasificación de los métodos iterativos (330).

Prólogo

La resolución numérica de las ecuaciones diferenciales de la física matemática por el método de las diferencias finitas se realiza en dos etapas: 1) la aproximación de diferencias de la ecuación diferencial sobre una red: escritura del esquema de diferencias, 2) la resolución en las CE de las ecuaciones en diferencias, las cuales representan sistemas de ecuaciones algebraicas lineales de alto orden y de un tipo especial (mal acondicionamiento, estructura de banda de la matriz del sistema). La aplicación de los métodos generales del álgebra lineal para tales sistemas no es siempre razonable tanto por la necesidad de conservar un gran volumen de información, como por el gran volumen de trabajo computacional, exigido por estos métodos. Para resolver ecuaciones en diferencias ya hace tiempo se elaboran métodos especiales, los cuales en uno u otro grado toman en consideración el carácter específico del problema y permiten hallar la solución con el gasto de un menor número de operaciones en comparación con los métodos generales del álgebra lineal.

Este libro es una continuación del libro de A.A. Samarski y V.B. Andreiev «Métodos de diferencias de resolución de ecuaciones elípticas», en el cual se estudia un círculo de problemas relacionados con la aproximación de diferencias, con la construcción de operadores de diferencias y con la estimación de la velocidad de convergencia de los esquemas de diferencias para los problemas de contorno típicos de tipo elíptico.

Aquí, nosotros examinamos sólo los métodos de resolución de ecuaciones en diferencias. El libro prácticamente consta de dos partes. La primera parte (cap. I-IV) está

dedicada a la aplicación de los métodos directos de resolución de las ecuaciones en diferencias y la segunda parte (cap. V-XV) a la teoría de los métodos iterativos de resolución de las ecuaciones reticulares de tipo general y su aplicación a las ecuaciones en diferencias. Al utilizar los métodos directos juega un papel esencial el tipo especial de las ecuaciones en diferencias. Para resolver las ecuaciones tripuntuales unidimensionales se examinan diferentes variantes del método de factorización (monótona, no monótona, cíclica, por flujos y otras).

En los capítulos III y IV se exponen los métodos directos económicos modernos para resolver las ecuaciones de Poisson en diferencias en un rectángulo con condiciones de contorno de diferente tipo. Estos son el método de reducción completa y el método de separación de variables, el cual utiliza el algoritmo de la transformación de Fourier rápida, y también los métodos combinados.

Durante el estudio de los métodos iterativos se utiliza la interpretación del método iterativo como un esquema operacional de diferencias, desarrollada en los libros de A.A. Samarski «Introducción a la teoría de los esquemas de diferencias» (1971) y «Teoría de los esquemas de diferencias» (1977). Esta concepción permite desarrollar la teoría de los métodos iterativos como una parte de la teoría general de estabilidad de los esquemas operacionales de diferencias, sin recurrir a suposiciones sobre la estructura de la matriz del sistema (véase también A.A. Samarski y A.V. Gulin «Estabilidad de los esquemas de diferencias» (1973)). La escritura de los esquemas iterativos en la forma canónica permite no solamente separar los operadores que responden por la convergencia de las iteraciones, sino también comparar diferentes métodos iterativos. La atención fundamental se presta al estudio de la velocidad de convergencia de las iteraciones y a la elección de los parámetros óptimos, para los cuales la velocidad de convergencia es máxima. La presencia de las estimaciones de la velocidad de convergencia, y también la investigación del carácter de la estabilidad computacional permiten realizar la comparación de diferentes métodos iterativos en situaciones concretas y hacer una elección. Aunque el lector, probablemente, conoce los fundamentos de la teoría de los esquemas de diferencias y los elementos del análisis funcional, sin embargo en el capítulo V se citan los conocimientos utilizados en el libro sobre los fundamentos del aparato matemático de la teoría

de los esquemas iterativos y se muestra, como las aproximaciones de diferencias de las ecuaciones elípticas se reducen a las ecuaciones operacionales de primer género $Au = f$ con operadores A en un espacio de Hilbert de funciones reticulares.

En los siguientes capítulos se investigan el esquema iterativo de dos capas con el conjunto de los parámetros de Chebishev, para lo cual tiene lugar la estabilidad computacional del método; el esquema de tres capas; los métodos iterativos de tipo variacional (métodos del descenso más rápido, de los defectos mínimos, de las correcciones mínimas, de los gradientes conjugados y otros); los métodos iterativos para ecuaciones no autoconjugadas y en el caso de un operador degenerado sin signo definido; los métodos de las direcciones variables; los métodos «triangulares» (con el algoritmo de inversión de la matriz triangular para la determinación de una nueva iteración) tales, como el método de Seidel, el método de relajación superior y otros; los métodos iterativos de resolución de ecuaciones en diferencias no lineales, la resolución de los problemas de diferencias de contorno para ecuaciones elípticas en sistemas curvilíneos de coordenadas y otros.

Un lugar especial en el libro lo ocupa el método universal alternativo triangular, propuesto y desarrollado por los autores en los años 1964—1977, cuya efectividad se manifiesta fuertemente durante la resolución del problema de Dirichlet para la ecuación de Poisson en una región arbitraria y del problema de Dirichlet para la ecuación div $(k \text{ grad } u) = -f(x)$, $x = (x_1, x_2)$ con un coeficiente $k(x)$ que varía fuertemente.

En el libro se muestra, como pasar de la teoría general a los problemas concretos, y se cita un gran número de algoritmos iterativos de resolución de ecuaciones en diferencias para ecuaciones elípticas y sistemas de ecuaciones. Son dadas estimaciones del número de iteraciones y es realizada una comparación de los diferentes métodos. Así, en particular, se muestra, que para el problema más simple los métodos directos son más económicos, que el método de las direcciones variables. Se debe subrayar, que todos los problemas cada vez más complejos del álgebra lineal que aparecen en la práctica exigen tanto la elaboración de nuevos métodos, como la extensión del dominio de aplicabilidad de los métodos viejos. Con esto ocurre una valoración de las características comparables de los diferentes métodos.

Durante la escritura del libro los autores utilizaron los materiales de las lecciones dictadas por ellos en el período de los años 1961—1977 en la facultad mecánico-matemática y en la facultad de matemática de cálculo y cibernética de la Universidad de Moscú Lomonosov, y también los materiales de los trabajos publicados de los autores.

Los autores aprovechan la oportunidad para expresar su agradecimiento a V.B. Andreiev, I.V. Friazinov, M.I. Bakirova, A.B. Kúcherov y I.E. Kaporin por la serie de útiles observaciones con respecto al material del libro.

Los autores están agradecidos a T.N. Galishnikova, A.A. Gólubeva y especialmente a V.M. Márchenko por la ayuda durante la preparación de manuscrito para su impresión.

A.A. Samarski, E.S. Nikolaiev.

Moscú, diciembre 1977.

Introducción

La aplicación de distintos métodos numéricos (de los métodos de diferencias, variacionales en diferencias, de proyecciones en diferencias e incluso del método de los elementos finitos) para resolver las ecuaciones diferenciales conduce a un sistema de ecuaciones algebraicas lineales de un tipo especial: a las ecuaciones en diferencias. Dicho sistema posee los siguientes rasgos específicos: 1) posee un orden alto, igual al número de nodos de la red; 2) el sistema está mal condicionado (la relación entre el valor propio máximo y el mínimo de la matriz del sistema es grande; así por ejemplo para el operador de Laplace en diferencias esta relación es inversamente proporcional al cuadrado del paso de la red); 3) la matriz del sistema está enrarecida, es decir, en cada una de sus filas son distintos de cero varios elementos cuyo número no depende de la cantidad de nodos; 4) los elementos no nulos de la matriz están distribuidos de una forma especial — la matriz es de banda.

Al aproximar las ecuaciones integrales e integro-diferenciales en la red, obtenemos un sistema de ecuaciones con respecto a una función definida sobre la red (función reticular). Es natural llamar ecuaciones reticulares a tales ecuaciones:

$$\sum_{\xi\in\omega} a(x, \xi)\, y(\xi) = f(x), \quad x \in \omega, \tag{1}$$

donde la sumación se efectúa según todos los nodos de la red ω, es decir, según un conjunto discreto de puntos. En el caso general, la matriz $(a(x, \xi))$ de la ecuación reticular está saturada. Si renumeramos los nodos de la red, entonces

la ecuación reticular se puede escribir en la forma

$$\sum_{j=1}^{N} a_{ij} y_j = f_i, \quad i = 1, 2, \ldots, N, \tag{2}$$

donde i, j son números de los nodos de la red y N es la cantidad total de nodos. El razonamiento inverso es evidente. De esta forma, una ecuación lineal reticular es un sistema de ecuaciones algebraicas lineales y recíprocamente todo sistema lineal de ecuaciones algebraicas puede ser interpretado como una ecuación reticular lineal respecto a una función reticular definida sobre cierta red, con un número de nodos igual al orden del sistema. Observaremos, que los métodos variacionales (de Ritz, de Galerkin y otros) de solución numérica de ecuaciones diferenciales conducen frecuentemente a sistemas con matriz saturada.

Una ecuación en diferencias es el caso particular de ecuación reticular, cuando la matriz (a_{ij}) está enrarecida. Así, por ejemplo, (2) representa una ecuación en diferencias de m-ésimo orden, si en la fila de número i se tienen solamente $m + 1$ elementos a_{ij} distintos de cero para $j = i, i + 1, \ldots$ $\ldots, i + m$.

De lo dicho, resulta claro que la solución de las ecuaciones reticulares y, en particular, de las ecuaciones en diferencias es un problema de álgebra lineal.

* * *

Para la solución de los problemas del álgebra lineal existen muchos métodos numéricos distintos, continuamente se trabaja en su perfeccionamiento y se realiza una revaloración de los métodos, elaborándose otros nuevos. Como conclusión resulta que un parte considerable de los métodos con que se cuenta tiene derecho a existir ya que posee su propio dominio de aplicabilidad. Por eso para la solución de un problema concreto en una ordenadora (computadora) existe el problema de la elección de un método en el conjunto de los métodos admisibles de solución de dicho problema. Evidentemente, este método debe poseer las mejores características posibles (o, como gustan decir, ser un método optimal) tales como el mínimo de tiempo de solución del problema en una ordenadora (o el mínimo del número de operaciones aritméticas y lógicas durante la búsqueda de la

solución), estabilidad computacional, es decir, estabilidad con respecto a los errores de redondeamiento y otros.

Es natural exigir, que cualquier algoritmo computacional para ordenadora permita en principio obtener la solución de un problema dado con cualquier exactitud $\varepsilon > 0$ prefijada con anterioridad, en un número finito de operaciones $Q(\varepsilon)$. Esta exigencia es satisfecha por un conjunto innumorable de algoritmos, entre los cuales se debe buscar el algoritmo con un mínimo $Q(\varepsilon)$ para todo $\varepsilon > 0$. Tal algoritmo se llama económico. Es claro que la búsqueda de un método «optimal» o «mejor posible» se efectúa en un conjunto de métodos conocidos (y no de todos los admisibles) y el mismo término de «algoritmo optimal» posee un sentido acotado y condicional.

* * *

El problema de la teoría de los métodos numéricos consiste tanto en la búsqueda de mejores algoritmos para una clase dada de problemas, como en el establecimiento de una jerarquía de métodos. El mismo concepto de mejor algoritmo depende del objetivo de los cálculos.

Son posibles dos planteamientos del problema acerca de la elección del mejor método:

a) se exige resolver un sistema concreto de ecuaciones $Au = f$, donde $A = (a_{ij})$ es una matriz;

b) se exige resolver algunas variantes de un mismo problema, por ejemplo, de la ecuación $Au = f$ con distintos miembros segundos f.

En un cálculo multivariante se puede disminuir el número medio de operaciones $\bar{Q}(\varepsilon)$ para una variante, si se conservan ciertas magnitudes y no se calculan de nuevo para cada variante (por ejemplo, si se conserva la matriz inversa).

De aquí está claro que la elección de un algoritmo debe depender del tipo de cálculo (univariante o multivariante) y de la posibilidad de conservar información complementaria en la memoria de computadora, lo cual a su vez está relacionado tanto con el tipo de la computadora, como con el orden del sistema de ecuaciones. En las estimaciones teóricas de la calidad de un algoritmo computacional generalmente se limitan al recuento del número de operaciones aritméticas que se exigen para encontrar la solución con una exactitud prefijada mientras que, por regla general, no se

examina el problema sobre los parámetros de la computadora.

Los últimos años el desarrollo impetuoso de métodos numéricos de solución de ecuaciones en diferencias, que aproximan ecuaciones diferenciales de tipo elíptico, y la aparición de nuevos algoritmos económicos han conducido a la necesidad de revisar las ideas que se tenían sobre los dominios de aplicabilidad de los métodos antes existentes.

* * *

El contenido de este libro está determinado en un grado significativo por la necesidad de dar métodos efectivos de solución de las ecuaciones en diferencias, correspondientes a los problemas de contorno para ecuaciones de tipo elíptico de segundo orden. La clasificación de los problemas de contorno en diferencias puede ser realizada según los siguientes indicios:

1) el tipo del operador diferencial L en la ecuación:

$$Lu = f(x), \quad x = (x_1, x_2, \ldots, x_p) \in G; \tag{3}$$

2) la forma de la región G en la cual se busca la solución;

3) el tipo de las condiciones de contorno en la frontera Γ de la región G;

4) la red $\bar{\omega}$ en la región $\bar{G} = G + \Gamma$ y el esquema de diferencias

$$\Lambda y = -\varphi(x), \quad x \in \omega, \tag{4}$$

es decir, el tipo del operador de diferencias Λ.

Ejemplos de operador elíptico de segundo orden pueden ser

$$Lu = \Delta u = \sum_{\alpha=1}^{p} \frac{\partial^2 u}{\partial x_\alpha^2} \text{ — operador de Laplace}, \tag{5}$$

$$Lu = \sum_{\alpha,\beta=1}^{p} \frac{\partial}{\partial x_\alpha}\left(k_{\alpha\beta}(x)\frac{\partial u}{\partial x_\beta}\right) - q(x)u, \tag{6}$$

además los coeficientes $k_{\alpha\beta}(x)$ satisfacen en cada punto $x = (x_1, x_2, \ldots, x_p)$ la condición de elipticidad fuerte

$$c_1 \sum_{\alpha=1}^{p} \xi_\alpha^2 \leqslant \sum_{\alpha,\beta=1}^{p} k_{\alpha\beta}(x)\xi_\alpha\xi_\beta \leqslant c_2 \sum_{\alpha=1}^{p} \xi_\alpha^2, \quad c_1, c_2 = \text{const} > 0, \tag{7}$$

donde $\xi = (\xi_1, \ldots, \xi_p)$ es un vector arbitrario. Si $u(x) = (u^1(x), u^2(x), \ldots, u^m(x))$ es una función vectorial, entonces (3) es un sistema de ecuaciones y

$$(Lu)^i = \sum_{j=1}^{m} \sum_{\alpha, \beta=1}^{p} \frac{\partial}{\partial x_\alpha} \left(k_{\alpha\beta}^{ij} \frac{\partial u^j}{\partial x_\beta} \right), \quad i = 1, 2, \ldots, m,$$

y la condición de elipticidad fuerte tiene la forma

$$c_1 \sum_{i=1}^{m} \sum_{\alpha=1}^{p} (\xi_\alpha^i)^2 \leqslant \sum_{i, j=1}^{m} \sum_{\alpha, \beta=1}^{p} k_{\alpha\beta}^{ij}(x) \xi_\alpha^i \xi_\beta^j \leqslant c_2 \sum_{i=1}^{p} \sum_{\alpha=1}^{m} (\xi_\alpha^i)^2.$$

$c_1, c_2 = \text{const} > 0$.

* * *

La forma de la región influye fuertemente en las propiedades de la matriz de las ecuaciones en diferencias. Nosotros distinguiremos regiones para las cuales la ecuación $Lu = 0$ con condiciones de contorno homogéneas admite separación de variables. Así, por ejemplo, para la ecuación de Laplace en coordenadas cartesianas (x_1, x_2), $Lu = \Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2}$, el método de separación de variables es aplicable en el caso, cuando G es un rectángulo. Una propiedad análoga posee un esquema de diferencias sobre una red rectangular, por ejemplo, el esquema «cruz»; en este caso la red puede ser no uniforme por cada dirección.

Al comparar distintos métodos numéricos de solución de sistemas de ecuaciones algebraicas utilizaremos en calidad de problema *patrón ó modelo*, el siguiente problema de contorno de diferencias:

ecuación de Poisson, dominio —un cuadrado, condiciones de contorno de primer género, red— cuadrada con pasos $h_1 = h$ y $h_2 = h$ según x_1 y x_2, operador de diferencias Λ-pentapuntual.

El segundo grupo de problemas de contorno de diferencias corresponde a los siguientes datos: L — operador con coeficientes variables del tipo (6): a) sin derivadas mixtas, b) con derivadas mixtas, la región $G = \{0 \leqslant x_\alpha \leqslant l_\alpha, \alpha = 1, 2\}$ — un rectángulo (paralelepípedo cuando $p \geqslant 3$).

El tercer grupo de problemas es la región de forma compleja, y L es el operador de Laplace o un operador de tipo general. Aquí el grado de complejidad del problema se determina en primer lugar por la forma de la región, la elección

de la retícula y la elección del operador de diferencias en una vecindad de la frontera.

Para el segundo y tercer grupo de problemas el operador de diferencias se elige frecuentemente de manera tal que se conserven las propiedades fundamentales (autoadjunticidad, definición del signo y otras) del problema inicial y se satisfagan las exigencias de aproximación con un orden determinado respecto al paso de la retícula.

* * *

Para resolver los problemas elípticos de diferencias se aplican métodos directos e iterativos.

Los métodos directos son aplicables en el caso multidimensional fundamentalmente para problemas del primer grupo (L — operador de Laplace, G — un rectángulo siendo $p = 2$ y un paralelepípedo si $p \geqslant 3$ y Λ es un esquema de diferencias pentapuntual o nonapuntual cuando $p = 2$). Para problemas unidimensionales, cuando la ecuación en diferencias tiene segundo orden (la matriz es tridiagonal), y los coeficientes de la ecuación pueden ser variables, aplicaremos el método de factorización que es una variante del método de Gauss (véase cap. II). Existe una serie de variantes del metodo de factorización: factorización monótona, factorización no monótona, factorización por flujos, factorización cíclica y otras (véase cap. II). Para los problemas bidimensionales del primer grupo (véase más arriba) resulta efectivo el método de reducción completa (cap. III), el método de separación de variables con la transformación de Fourier rápida, y también una combinación del método de reducción incompleta con la transformación de Fourier rápida (cap. IV). En todos los casos se resuelve una ecuación en diferencias de segundo orden por una de las direcciones mediante el método de factorización.

Los métodos directos señalados, en el caso del problema de diferencias de Dirichlet para la ecuación de Poisson en el rectángulo $(0 \leqslant x_\alpha \leqslant l_\alpha,\ \alpha = 1, 2)$ sobre la retícula $\bar{\omega} = \{(i_1h_1, i_2h_2),\ i_\alpha = 0, 1, \ldots, N_\alpha,\ h_\alpha = l_\alpha/N_\alpha,\ \alpha = 1, 2\}$, exigen $Q = O(N_1N_2 \log_2 N_2)$ operaciones aritméticas, donde $N_2 = 2^n$, siendo $n > 0$ un número entero.

Los métodos directos son aplicables a una clase muy especial de problemas.

* * *

Los problemas elípticos de diferencias, en el caso de operadores L de tipo general o de regiones de forma complicada, se resuelven fundamentalmente con ayuda de métodos iterativos.

Las ecuaciones reticulares se pueden interpretar como ecuaciones operacionales de primer género

$$Au = f \tag{8}$$

con operadores definidos sobre espacios H de funciones reticulares. En el espacio H se introducen un producto escalar (,) y normas energéticas $\| u \|_D = \sqrt{(Du, u)}$, $D = D^* > 0$, D: $H \to H$, donde D es un cierto operador lineal en H.

Los métodos iterativos de solución de la ecuación operacional $Au = f$ pueden ser interpretados como ecuaciones operacionales en diferencias (o de diferencias respecto a un tiempo ficticio o al número-índice de la iteración) con operadores en el espacio de Hilbert H. Si una nueva iteración y_{k+1} se calcula mediante las m iteraciones anteriores y_k, $y_{k-1}, \ldots, y_{k-m+1}$, entonces el método iteracional (esquema) se llama $m+1$ capas (de m pasos). De aquí se ve la analogía entre los esquemas iterativos y los esquemas de diferencias para problemas no estacionarios. Por eso la teoría de los métodos iterativos prácticamente es una parte especial de la teoría general de estabilidad de los esquemas operacionales de diferencias. Nosotros nos limitaremos al estudio de los esquemas de dos capas y en menor grado de los de tres capas. El paso a los esquemas de capas múltiples no proporciona ninguna ventaja (como se deduce de la teoría general de estabilidad, véase [10]).

Un importante papel juega la notación de los métodos iterativos en una forma única (canónica), lo cual permite distinguir el operador (estabilizador) que responde por la estabilidad y convergencia de las iteraciones y comparar métodos iterativos distintos desde una misma posición.

Cualquier método iteracional de dos capas (de un paso) se escribe en la siguiente forma canónica:

$$B\,\frac{y_{k+1}-y_k}{\tau_{k+1}} + Ay_k = f \quad k = 0,\ 1,\ \ldots,\ y_0 \in H, \tag{9}$$

donde B: $H \to H$ es un operador lineal que posee inverso B^{-1}; $\tau_1, \tau_2, \ldots$ son los parámetros iterativos; k es el número de iteración y y_k es la aproximación iterativa del número k.

En el caso general $B = B_{k+1}$ depende de k. En la teoría general nosotros suponemos que B no depende de k.

Los parámetros $\{\tau_k\}$ y el operador B son arbitrarios y hay que elegirlos de la condición de mínimo de iteraciones n, para el cual la solución y_n de la ecuación (9) aproxima en H_D la solución exacta u de la ecuación $Au = f$ con una exactitud relativa $\varepsilon > 0$:

$$\| y_n - u \|_D \leqslant \varepsilon \| y_0 - u \|_D. \tag{10}$$

Para la teoría general de los métodos iterativos expuesta en este libro no se exige ninguna suposición sobre la estructura del operador A (de la matriz (a_{ij})). Se utilizan solamente propiedades de tipo general

$$\begin{gathered} A = A^* > 0, \quad B = B^* > 0, \\ \gamma_1 B \leqslant A \leqslant \gamma_2 B, \quad \gamma_1 > 0. \end{gathered} \tag{11}$$

Las desigualdades operacionales significan que están prefijadas las constantes γ_1 y γ_2 de equivalencia energética de los operadores A y B o los límites del espectro del operador A en el espacio H_B (γ_1 y γ_2 son los valores propios mínimo y máximo respectivamente del problema generalizado por valores propios: $Av = \lambda Bv$).

* * *

La solución $\tau_1, \tau_2, \ldots, \tau_n$ del problema indicado más arriba sobre el $\min\limits_{\tau_1, \tau_2, \ldots, \tau_n} n_0(\varepsilon)$ para γ_1 y γ_2 dados y B fijo, para el caso $D = AB^{-1}A$ se expresa mediante los ceros del polinomio de Chébishev de n-ésimo orden (método iteracional de Chébishev). Para estos valores óptimos $\tau_1, \tau_2, \ldots, \tau_n$ y prefijado $\varepsilon > 0$ arbitrario, es válida la estimación

$$n \geqslant \frac{\ln(2/\varepsilon)}{\ln\left((1+\sqrt{\xi})/(1-\sqrt{\xi})\right)},$$

ó

$$n \geqslant n_0(\varepsilon) = \frac{\ln(2/\varepsilon)}{2\sqrt{\xi}}, \quad \xi = \gamma_1/\gamma_2,$$

para el número de iteraciones n que se calculan según el esquema (9) y se cumple la desigualdad

$$\| Ay_n - f \|_{B^{-1}} \leqslant \varepsilon \| Ay_0 - f \|_{B^{-1}}.$$

La estabilidad computacional del método de Chébishov tiene lugar para un modo determinado de numeración (ordenamiento) de los ceros del polinomio de Chébishov y de los parámetros $\tau_1^*, \tau_2^*, \ldots, \tau_n^*$; este método se muestra en el cap. VI.

Cuando $B = E$ (E — operador unidad) el método (9) se llama explícito y cuando $B \neq E$ se llama implícito. Si elegimos el parámetro τ_k constante, $\tau_k = \tau_0 = 2/(\gamma_1 + \gamma_2)$, $k = 1, 2, \ldots, n$, entonces obtenemos un esquema implícito de iteración simple para el cual $n \geqslant n_0(\varepsilon) = \ln\left(\frac{1}{\varepsilon}\right)\Big/(2\xi)$.

El operador B (estabilizador) se elige de la condición de economicidad, es decir, del mínimo de trabajo computacional al resolver la ecuación $Bv = F$, siendo prefijado el miembro F y, como ha dicho, de la condición de mínimo del número de iteraciones $n_0(\varepsilon)$.

Supongamos que nosotros sabemos resolver de modo económico el problema $Rv = f$ con un gasto $Q_R(\varepsilon)$ de operaciones, donde

$$R: H \to H, \quad R = R^* > 0, \quad c_1 R \leqslant A \leqslant c_2 R, \quad c_1 > 0. \tag{12}$$

Entonces se puede poner que $B = R$ y hallar la solución del problema $Au = f$ mediante el esquema (9) con los parámetros $\{\tau_k^*\}$ si $\gamma_1 = c_1$, $\gamma_2 = c_2$, efectuando $Q_A(\varepsilon) \approx \frac{1}{2}\sqrt{c_2/c_1} \times$ $\times \ln(2/\varepsilon)\, Q_R(\varepsilon)$ operaciones.

Si, por ejemplo, L es un operador de tipo general y G es un rectángulo, entonces en calidad de R se puede tomar el operador de diferencias de Laplace pentapuntual y resolver la ecuación $Rv = f$ por el método directo.

Puede resultar, que sea más ventajoso resolver la ecuación $Rv = f$, no por el método directo sino por un método iterativo. Entonces $B \neq R$ y no se escribe en forma explícita, sino se realiza como resultado del procedimiento iterativo.

* * *

Los conocidos métodos de Zeidel y de relajación superior son implícitos y corresponden a matrices triangulares (operadores) B. La convergencia de estos métodos se demuestra a base de la teoría general de esquemas de diferencias (véase A.A. Samarski, Teoría de esquemas de diferencias, Moscú,

1977, en ruso ó A.A. Samarski, A. V. Gulin, Estabilidad de esquemas de diferencias, Moscú, 1973, en ruso). Sin embargo, para los métodos de Zeidel y de relajación superior el operador B no es autoconjugado y por eso no se puede utilizar el método de Chebishev (9) con un conjunto optimal de parámetros iterativos $\tau_1^*, \tau_2^*, \ldots, \tau_n^*$, lo cual permitiría aumentar la velocidad de convergencia de las iteraciones. El operador B se puede hacer autoconjugado, si se pone igual al producto de operadores conjugados uno del otro:

$$B = (E + \omega R_1)(E + \omega R_2), \quad R_2^* = R_1, \tag{13}$$

donde $\omega > 0$ es un parámetro. En calidad de R_1 y R_2 se pueden tomar operadores que posean matrices triangulares (R_1, inferior y R_2, superior), de manera tal que $R_1 + R_2 = R$: $H \to H$ y $R^* = R > 0$. En particular se puede suponer que

$$R_1 + R_2 = A, \quad R_2^* = R_1. \tag{14}$$

Es típica la siguiente suposición:

$$R \geqslant \delta E, \quad R_1 R_2 \leqslant \frac{\Delta}{4} A, \quad \delta > 0, \quad \Delta > 0. \tag{15}$$

Eligiendo luego $\omega = 2/\sqrt{\delta\Delta}$ de la condición mín $n_0(\varepsilon)$, encontramos los parámetros γ_1 y γ_2 y calculamos los parámetros $\{\tau_k^*\}$. La definición de y_{k+1} por medio de y_k y f se reduce a la solución sucesiva de dos sistemas de ecuaciones con matrices triangulares superior e inferior.

El método iterativo (9) construido mediante el operador factorizado B del tipo (13) lo llamaremos método triangular alternativo (MTA). El MTA evidentemente es universal, ya que la representación de A en forma de una suma $R_1 + R_2 = A$, donde $R_2^* = R_1$, es siempre posible. En el caso de un problema elíptico de diferencias la construcción de R_1 y R_2 no presenta dificultad. Así, por ejemplo, si Ay es el operador de Laplace de diferencias de $2p + 1$ puntos, es decir, $Ay \to -\sum_{\alpha=1}^{p} y_{\bar{x}_\alpha x_\alpha}$, entonces $R_1 y \to \sum_{\alpha=1}^{p} \frac{y_{\bar{x}_\alpha}}{h_\alpha}$ y $R_2 y \to -\sum_{\alpha=1}^{p} \frac{y_{x_\alpha}}{h_\alpha}$, donde h_α es el paso de la retícula por la dirección Ox_α. Este método es rápidamente convergente. Si tomamos el conjunto de Chebishev $\{\tau_k^*\}$ y tenemos en cuenta (14) y (15), entonces el número de iteraciones para

el MTA será

$$n_0(\varepsilon) \geqslant \frac{1}{2\sqrt{2}\sqrt[4]{\eta}} \ln\frac{2}{\varepsilon}, \quad \eta = \frac{\delta}{\Delta}. \qquad (16)$$

En particular, para el problema modelo tenemos $n \geqslant$ $\geqslant n_0(\varepsilon) = 0{,}3 \ln\frac{2}{\varepsilon}\Big/\sqrt{h}$.

Para el caso de una región arbitraria y de ecuaciones con coeficientes variables, resulta útil emplear el método triangular alternativo modificado (MTAM), suponiendo que

$$B = (\mathscr{D} + \omega R_1)\mathscr{D}^{-1}(\mathscr{D} + \omega R_2), \quad R_2^* = R_1 \quad \mathscr{D} = \mathscr{D}^* > 0, \qquad (17)$$

donde $\mathscr{D}$ es un operador arbitrario. Si en lugar de (15) se cumplen

$$R \geqslant \delta\mathscr{D}, \quad R_1\mathscr{D}^{-1}R_2 \leqslant \frac{\Delta}{4}\mathscr{D}, \quad \delta > 0, \quad \Delta > 0, \qquad (18)$$

entonces la estimación (16) conserva su validez.

Aquí son prefijados δ y Δ y se eligen el operador $\mathscr{D}$ y el parámetro ω de manera tal que la relación $\xi = \gamma_1/\gamma_2$ sea máxima. En la práctica, en calidad de la matriz $\mathscr{D}$ se puede tomar una matriz diagonal.

Indiquemos dos ejemplos de aplicación efectiva del MTAM.

1) Problema de Dirichlet para la ecuación de Poisson en una región bidimensional de forma complicada; el retículo fundamental en el plano (x_1, x_2) es uniforme con paso h y el esquema es pentapuntual. Para la correspondiente elección de $\mathscr{D}$ el MTAM exige solamente un 4—5% más iteraciones que el mismo problema en el cuadrado con lado igual al diámetro de la región.

2) Para las ecuaciones elípticas con coeficientes que varían bruscamente (la relación c_2/c_1 es grande), el MTAM con $\mathscr{D}$ elegido de un modo apropiado permite debilitar la dependencia de c_2/c_1.

En la práctica además de los métodos de un paso (de dos capas) (9) se aplican los esquemas iterativos de dos pasos (de tres capas). Para parámetros iterativos optimales estos métodos son comparables, por el número de iteraciones, con el esquema de Chebishev de parámetros $\{\tau_k^*\}$, cuando $\xi \to 0$. Sin embargo, ellos son más sensibles a los errores en la definición de γ_1 y γ_2. Bajo las condiciones (11) es más racional utilizar el esquema de Chebishev (9) de los parámetros $\{\tau_k^*\}$.

* * *

Para la solución de problemas elípticos jugó un papel muy importante el método de las direcciones variables (MDV), desarrollado desde los principios de 1955 por muchos autores. Sin embargo, él resultó económico solamente para una clase muy estrecha de problemas del primer grupo, cuando se cumplen las condiciones $A = A_1 + A_2$, $A_\alpha = A_\alpha^* \geqslant 0$, $\alpha = 1, 2$, $A = A^* > 0$ y $A_1A_2 = A_2A_1$. Si A_1 y A_2 son conmutables, entonces para el MDV se pueden elegir parámetros iterativos optimales. Para el problema modelo con tales parámetros, el número de iteraciones $n_0(\varepsilon) = O\left(\ln \frac{1}{h} \ln \frac{1}{\varepsilon}\right)$ y el número de operaciones es $Q(\varepsilon) = O\left(\frac{1}{h^2} \ln \frac{1}{h} \ln \frac{1}{\varepsilon}\right)$ mientras que para los métodos directos $Q = O\left(\frac{1}{h^2} \ln \frac{1}{h}\right)$. En este caso los métodos directos son más económicos que el MDV. Si A_1 y A_2 son no conmutables, entonces el MDV exige $O\left(\frac{1}{h} \ln \frac{1}{\varepsilon}\right)$ iteraciones, mientras que para el MTA es suficiente $O\left(\frac{1}{\sqrt{h}} \ln \frac{1}{\varepsilon}\right)$ iteraciones. En el caso de los problemas tridimensionales, cuando $A = A_1 + A_2 + A_3$, aún bajo la suposición de conmutatividad de A_1, A_2 y A_3 dos a dos, el MDV exige más operaciones que el MTA. De esta forma el MDV ha perdido su valor en un grado significativo.

* * *

Si el operador $A > 0$ no es autoconjugado, entonces no tiene éxito construir mediante el esquema (9) con un conjunto de parámetros y un operador autoconjugado $B = B^* > 0$ un proceso iterativo de la misma velocidad de convergencia que el método de Chobishev para $A = A^* > 0$. Todos los métodos conocidos poseen una velocidad de convergencia menor. Aquí se examina el método de iteración simple (cap. VI) siendo prefijada una información de dos tipos:

a) se dan los parámetros γ_1 y γ_2 que entran en los datos (para simplificar consideraremos $D = B = E$).

$$\gamma_1 (x, x) \leqslant (Ax, x), \quad (Ax, Ax) \leqslant \gamma_2 (Ax, x),$$
$$\gamma_1 > 0, \quad \gamma_2 > 0; \tag{19}$$

b) se dan tres parámetros γ_1, γ_2 y γ_3, dondo γ_1 y γ_2 (para $D = B = E$), son las fronteras de la parte simétrica del operador A:

$$\gamma_1 E \leqslant A \leqslant \gamma_2 E, \quad \| A_1 \| \leqslant \gamma_3, \quad \gamma_1 > 0, \quad \gamma_3 \geqslant 0, \tag{20}$$

dondo $A_1 = 0{,}5\,(A - A^*)$ es la parte antisimétrica do A.

Eligiendo τ do la condición do mínimo de la norma del operador de transición o del operador do permisibilidad, en todos los casos obtenemos un aumento del número de iteraciones con respecto al caso $A = A^*$.

* * *

Todo método iterativo de dos capas, construido a base del esquema (9) se caracteriza por los operadores B y A, por el espacio energético H_D en el cual se demuestra la convergencia del método, y por el conjunto do parámetros. Si el operador B es fijo, entonces el problema fundamental es la búsqueda de los $\{\tau_k\}$.

Para la elección de los parámetros $\{\tau_k\}$ se utiliza la información previa sobre los operadores del esquema. El tipo de información se determina por las propiedades de los operadores A, B y D. Así, para el esquema de Chebishev si $D = AB^{-1}A$, cuando A y B son operadores autoconjugados, se supone que son prefijadas las constantes γ_1 y γ_2 en (11). En el caso general, cuando $DB^{-1}A$ es autoconjugado en H, entonces en lugar de (11) es suficiente exigir que $\gamma_1 D \leqslant \leqslant DB^{-1}A \leqslant \gamma_2 D$ con $\gamma_1 > 0$. En el caso no autoconjugado, cuando $A \neq A^*$, y $B = B^* > 0$, se utilizan dos números γ_1 y γ_2 o tres números γ_1, γ_2 (que entran en (19)) y la constante γ_3 que figura en la estimación de la parte antisimétrica del operador A. En una serie de casos la búsqueda de las constantes γ_1, γ_2 y γ_3 con suficiente exactitud, puede resultar un problema complejo independiente, que exija de algoritmos especiales para su solución. Si la información previa puede ser obtenida mediante un gasto computacional no muy grande o se exigen cálculos multivariantes para la ecuación $Au = f$ con distintos miembros derechos, entonces resulta útil hallar una vez los números necesitados γ_1, γ_2 y γ_3 y después utilizar el método de Chebishev o el método MTA. Si se exige resolver solamente un problema $Au = f$ o si se

da una buena aproximación inicial y el cálculo de las constantes γ_1 y γ_2 resulta laborioso, conviene utilizar métodos iterativos de tipo variacional.

Para los métodos iterativos de tipo variacional al calcular los parámetros $\{\tau_k\}$ no es necesario conocer γ_1 y γ_2. Estos métodos utilizan solamente información de tipo general

$$A = A^* > 0, \quad (DB^{-1}A)^* = DB^{-1}A. \qquad (21)$$

Para determinar y_{k+1} se emplea el mismo esquema (9), se cambia únicamente la fórmula para τ_{k+1}. El parámetro τ_{k+1} se halla de la condición de mínimo en H_D de la norma del error $z_{k+1} = y_{k+1} - u$, es decir, de mínimo del funcional $I\,[y] = (D\,(y - u),\ y - u)$. El parámetro τ_{k+1} se calcula mediante y_k. Eligiendo $D = A$, obtenemos el método de descenso más rápido, y para $D = A^*A$, el método de los defectos mínimos, etc. Estos métodos tienen la misma velocidad de convergencia que el método de iteración simple (para constantes exactas γ_1 y γ_2). La velocidad de convergencia de las iteraciones se puede aumentar si renunciamos a la minimización local (en cada paso) de $\| z_{k+1} \|_D$ y elegimos los parámetros τ_k de la condición de minimización de la norma del error $\| z_n \|_D$ inmediatamente después de n pasos, es decir, al pasar de y_0 a y_n. Este camino conduce a los esquemas iterativos biparamétricos (para cada k) de tres capas de direcciones conjugadas (de gradientes de defectos, de correcciones o de errores conjugados), los cuales poseen la misma velocidad de convergencia que el método de Chebishev con los parámetros $\{\tau_k^*\}$ calculados según los valores exactos de γ_1 y γ_2. Si $A = A^* > 0$, entonces se puede construir un proceso de aceleración ($\approx$ en 1,5—2 veces) de la convergencia para los métodos de dos capas gradientales.

* * *

En la teoría general de los métodos iterativos no se exige el conocimiento de la estructura concreta de los operadores del problema. Se utiliza solamente un mínimo de información de carácter funcional general respecto a los operadores, por ejemplo, la condición (11). La elección del operador B del esquema (9) está subordinada a las siguientes exigencias: 1) el aseguramiento de una convergencia más rápida del método (9), 2) el ahorro de la inversión de B. Al construir B se puede partir de cierto operador $R = R^* > 0$ (regulariza-

dor), energéticamente equivalente a $A = A^* > 0$, siendo $B = B^* > 0$:

$$c_1 R \leqslant A \leqslant c_2 R, \quad c_1 > 0, \quad \mathring{\gamma}_1 B \leqslant R \leqslant \mathring{\gamma}_2 B, \quad \mathring{\gamma}_1 > 0. \tag{22}$$

Por lo tanto, $\gamma_1 = c_1 \mathring{\gamma}_1$ y $\gamma_2 = c_2 \mathring{\gamma}_2$. Para distintos A se puede elegir un mismo regularizador R. El caso más difundido es el de un operador B factorizado, por ejemplo:

$$B = (E + \omega R_1)(E + \omega R_2), \quad R_1 + R_2 = R, \tag{23}$$

donde

$$R_1^* = R_2 > 0 \quad \text{para el MTA}, \tag{24}$$

$$R_1^* = R_1 > 0, \; R_2^* = R_2 > 0, \; R_1 R_2 = R_2 R_1 \text{ para el MDV}. \tag{25}$$

Para aplicar la teoría es necesario hallar $\mathring{\gamma}_1$ y $\mathring{\gamma}_2$; el parámetro $\omega > 0$ se encuentra de la condición mín $(\mathring{\gamma}_1(\omega)/\mathring{\gamma}_2(\omega))$. Si la ecuación $Rw = F$ puede ser resuelta por un método directo económico, entonces tomaremos $B = R$ (por ejemplo, en el caso cuando $(-R)$ es el operador de Laplace de diferencias y la región es un rectángulo). El operador B puede no escribirse explícitamente sino realizarse como resultado de la solución iterativa de la ecuación $Rw = r_k$, donde $r_k = A_{u_k} - f$ (método biescalonado).

* * *

Para ecuaciones con los operadores A sin signo definido, degenerados y complejos, se pueden examinar los mismos esquemas (9). Sin embargo, la elección de parámetros óptimales se complica, y la velocidad de convergencia de las iteraciones disminuye. La aplicación de la teoría general para estos casos singulares exige un «tratamiento» previo del problema inicial. Resulta posible construir una modificación tanto del método de Chebishev, como de los métodos de tipo variacional.

Si A es un operador lineal degenerado, es decir, la ecuación homogénea $Au = 0$ tiene una solución no trivial, entonces el problema (9) para $B = E$ y toda τ_k es siempre resoluble. Sea $H^{(0)}$ el subespacio propio nulo del operador A y $H^{(1)}$ es el complemento ortogonal de $H^{(0)}$ hasta H. Cualquier

vector $y \in H^{(0)}$ satisface la ecuación $Ay = 0$. Si $f \in H^{(1)}$ y $y_0 \in H^{(1)}$, entonces todas las iteraciones $y_k \in H^{(1)}$. Si se cumplen las condiciones

$$\gamma_1 (y, y) \leqslant (Ay, y) \leqslant \gamma_2 (y, y), \quad y \in H^{(1)}, \ \gamma_1 > 0,$$

entonces se puede utilizar el esquema explícito (9) con los parámetros de Chebishov $\{\tau_k^*\}$ hallados a base de γ_1 y γ_2. Con eso y_k converge a la solución normal, que posee norma mínima.

Si $f = f^{(0)} + f^{(1)}$ y $f^{(0)} \neq 0$, entonces entenderemos por solución normal generalizada de la ecuación $Au = f$, la solución de la ecuación $Au^{(1)} = f^{(1)}$, con $u^{(1)} \in H^{(1)}$ y que posee norma mínima. Es válida la estimación

$$\| y_n - u^{(1)} \| \leqslant \tilde{q}_n \| y_0 - u^{(1)} \|,$$

$$\tilde{q}_n = q_{n-1} (1 + (n-1)) \sqrt{\frac{1 - q_{n-1}^2}{\xi}},$$

$$q_n = \frac{2\rho_1^n}{1 + \rho_1^{2n}}, \quad \rho_1 = \frac{1 - \sqrt{\xi}}{1 + \sqrt{\xi}}, \quad \xi = \frac{\gamma_1}{\gamma_2}, \quad y_n, y_0 \in H^{(1)},$$

si $\tau_1^*, \tau_2^*, \ldots, \tau_{n-1}^*$ son los parámetros de Chebishov, y $\tau_n^* = -\sum_{j=1}^{n-1} \tau_j^*$. La velocidad de convergencia disminuye en comparación con el caso de A no degenerado para los mismos γ_1 y γ_2. Paralelamente a la modificación indicada del método de Chebishov son posibles también los métodos de tipo variacional.

La teoría general permite investigar un esquema implícito de iteración simple para el caso, cuando H es un espacio de Hilbert complejo, $A = \tilde{A} + qE$, $\tilde{A}$ es un operador hermítico, $q = q_1 + iq_2$ es un número complejo y además permite elegir el valor óptimo del parámetro iterativo. La transición al método de las direcciones variables tampoco presenta dificultad.

* * *

No es difícil utilizar los resultados de la teoría general para la solución de ecuaciones en diferencias que aproximan problemas de contorno para ecuaciones de tipo elíptico. En este caso es fácil formular las reglas generales de solución de los problemas de diferencias. Sea dada la ecuación en diferencias $Au = f$, donde $A: H \to H$ es un operador de

diferencias definido en el espacio H de las funciones reticulares prefijadas sobre la red ω. Al principio se estudian las propiedades generales del operador A y se establece, por ejemplo, su autoadjunticidad y positividad, $A = A^* > 0$, después se construye el operador $B = B^* > 0$ y se calculan las constantes γ_1, γ_2 y finalmente se encuentran $n = n_0(\varepsilon)$ y los parámetros $\{\tau_k^*\}$.

Si se trata del MTA con el operador factorizado $B = (\mathscr{D} + \omega R_1)\mathscr{D}^{-1}(\mathscr{D} + \omega R_2)$, entonces es necesario elegir la matriz $\mathscr{D}$ y las constantes δ y Δ (véase el cap. X) y conociendo δ y Δ se determinan ω, γ_1, γ_2, etc.

En el libro se citan muchos ejemplos de aplicación de los métodos directos e iterativos para resolver ecuaciones concretas de diferencias. En particular, en el capítulo XV se examinan métodos de solución de ecuaciones elípticas de diferencias en coordenadas curvilíneas: en el sistema de coordenadas cilíndricas (r, z) y polares (r, φ).

En el capítulo XIV se examinan problemas multidimensionales, esquemas para las ecuaciones de la teoría de elasticidad y otros.

Es importante señalar, que independientemente del método que sea aplicado para resolver un problema de contorno de diferencias dado, su tratamiento previo se realiza por una misma receta: primeramente se forma el operador A y después se estudia como operador en un espacio H de funciones reticulares. Después que la «recolección» de información acerca del problema ha terminado, entonces se toma una decisión para elegir el método de solución del problema teniendo en cuenta todas las circunstancias, incluso el tipo de computadora, la existencia de programas estandarizados y otras.

Capítulo

Métodos directos de solución de ecuaciones en diferencias

En este capítulo se estudia la teoría general de las ecuaciones lineales en diferencias y además los métodos directos de solución de ecuaciones con coeficientes constantes, los cuales dan la solución en forma cerrada. En el § 1 se citan los conceptos generales sobre las ecuaciones reticulares. El § 2 está dedicado a la teoría general de ecuaciones lineales en diferencias de m-ésimo orden. En el § 3 se examinan los métodos de solución de ecuaciones con coeficientes constantes y en el § 4 se utilizan estos métodos para resolver ecuaciones de segundo orden. El § 5 se dedica a la solución de problemas reticulares en valores propios para el operador en diferencias más sencillo.

§ 1. Ecuaciones reticulares. Conceptos fundamentales

1. Retículos y funciones reticulares. Un número significativo de problemas de la física y la técnica se reduce a las ecuaciones diferenciales de derivadas parciales (a las ecuaciones de la física matemática). Los procesos estacionarios de distinta naturaleza física se describen mediante ecuaciones de tipo elíptico.

Las soluciones exactas de los problemas de contorno para ecuaciones elípticas se logran obtener solamente en casos particulares. Por eso, en general, estos problemas se resuelven aproximadamente. Uno de los métodos más universales y efectivos que ha obtenido en la actualidad una amplia difusión para la solución aproximada de las ecuaciones de la física matemática, es el método de diferencias finitas o método de retículos.

La idea del método consiste en lo siguiente. La región donde cambian continuamente los argumentos (por ejemplo, un segmento, un rectángulo, etc.) se cambia por un conjunto discreto de puntos (nodos) el cual se llama *red* o *retículo.*

En lugar de funciones de argumento continuo se examinan funciones de argumento discreto, llamadas *funciones reticulares* y que están definidas en los nodos del retículo.

Las derivadas que entran en la ecuación diferencial y en las condiciones de contorno se sustituyen por derivadas de diferencias. En este caso el problema de contorno para la ecuación diferencial se cambia por un sistema de ecuaciones algebraicas lineales o no lineales (ecuaciones reticulares o en diferencias). Tales sistemas se les llaman frecuentemente *esquemas de diferencias.*

Detengámonos más detalladamente en los conceptos fundamentales del método de retículos. Examinemos primeramente los ejemplos más sencillos de retículos.

EJEMPLO 1. *Retículos en una región unidimensional.* Sea el segmento $0 \leqslant x \leqslant l$, la región de cambio del argumento x. Dividamos este segmento en N partes iguales de longitud $h = l/N$, por los puntos $x_i = ih$, $i = 0, 1, \ldots, N$. El conjunto de estos puntos se llama *retículo uniforme* en el segmento $[0, l]$ y se denota por $\bar{\omega} = \{x_i = ih,\ i = 0, 1, \ldots, N,\ hN = l\}$, y el número h que es la distancia entre los puntos (nodos), del retículo $\bar{\omega}$, se le llama *paso* de la red.

Para separar una parte de la red $\bar{\omega}$, nosotros utilizaremos en lo que sigue la siguiente notación:

$$\omega = \{x_i = ih, \quad i = 1, 2, \ldots, N-1, \quad Nh = l\},$$
$$\omega^+ = \{x_i = ih, \quad i = 1, 2, \ldots, N, \quad Nh = l\},$$
$$\omega^- = \{x_i = ih, \quad i = 0, 1, \ldots, N-1, \quad Nh = l\},$$
$$\gamma = \{x_0 = 0, \quad x_N = l\}.$$

El segmento $[0, l]$ puede ser dividido en N partes, introduciendo puntos arbitrarios $0 = x_0 < x_1 < \ldots < x_i < x_{i+1} < \ldots \leqslant x_{N-1} < x_N = l$. En este caso obtendremos el retículo $\bar{\omega} = \{x_i,\ i = 0, 1, \ldots, N,\ x_0 = 0,\ x_N = l\}$ con paso $h_i = x_i - x_{i-1}$ en el nodo x_i, $i = 1, 2, \ldots, N$, el cual depende del número i del nodo x_i, es decir, es una función reticular $h_i = h(i)$.

Si $h_i \neq h_{i+1}$ al menos para un número i, entonces el retículo $\bar{\omega}$ se llama *no uniforme.* Si $h_i = h = l/N$, entonces obtendremos el retículo uniforme construido más arriba. Para el retículo no uniforme se introduce el paso medio $\hbar_i = \hbar(i)$ en el nodo x_i, como $\hbar_i = 0{,}5\,(h_i + h_{i+1})$, $1 \leqslant i \leqslant N-1$, donde $\hbar_0 = 0{,}5h_1$ y $\hbar_N = 0{,}5h_N$. Sobre la recta infinita $-\infty < x < \infty$ se pueden examinar los

retículos $\Omega = \{x_i = a + ih, \ i = 0, \ \pm 1, \ \pm 2, \ \ldots\}$ con inicio en cualquier punto $x = a$ y con paso h, compuestos por un número infinito de nodos.

EJEMPLO 2. *Retículo en una región bidimensional.* Sea la región de variación de los argumentos $x = (x_1, x_2)$ el rectángulo $\overline{G} = \{0 \leqslant x_\alpha \leqslant l_\alpha, \ \alpha = 1, 2\}$ con frontera Γ. En los segmentos $0 \leqslant x_\alpha \leqslant l_\alpha$ construiremos retículos uniformes $\overline{\omega}_\alpha$ con pasos h_α:

$$\overline{\omega}_1 = \{x_1(i) = ih_1, \quad i = 0, 1, \ldots, M, \quad h_1 M = l_1\},$$
$$\overline{\omega}_2 = \{x_2(j) = jh_2, \quad j = 0, 1, \ldots, N, \quad h_2 N = l_2\}.$$

El conjunto de nodos $x_{ij} = (x_1(i), x_2(j))$, que poseen coordenadas en el plano $x_1(i)$ y $x_2(j)$, se le llama retículo en el rectángulo $\overline{G}$ y se denota por $\overline{\omega} = \{x_{ij} = (ih_1, jh_2), i = 0, 1, \ldots, M, j = 0, 1, \ldots, N, h_1 M = l_1, h_2 N = l_2\}$.

Evidentemente, el retículo $\overline{\omega}$ está compuesto por los puntos de intersección de las rectas $x_1 = x_1(i)$ y $x_2 = x_2(j)$.

El retículo $\overline{\omega}$ construido es uniforme respecto de cada una de las variables x_1 y x_2. Si al menos uno de los retículos $\overline{\omega}_\alpha$ no es uniforme entonces el retículo $\overline{\omega}$ se llama *no uniforme*. Si $h_1 = h_2$, entonces el retículo se llama *cuadrado* ó *rectangular*.

Los puntos de $\overline{\omega}$ pertenecientes a Γ, se llaman puntos de *frontera* y su unión constituye la frontera de la red: $\gamma = \{x_{ij} \in \Gamma\}$.

Para describir la estructura del retículo $\overline{\omega}$ es cómodo utilizar la escritura $\overline{\omega} = \overline{\omega}_1 \times \overline{\omega}_2$, es decir, representarse $\overline{\omega}$ como el producto topológico de los retículos $\overline{\omega}_1$ y $\overline{\omega}_2$.

Empleando las notaciones introducidas en el ejemplo 1 de ω^+, ω^- y ω, se pueden separar partes del retículo $\overline{\omega}$ en el rectángulo, por ejemplo:

$$\omega_1 \times \omega_2^+ = \{x_{ij} = (ih_1, jh_2), i = 1, 2, \ldots, M-1, \\ j = 1, 2, \ldots, N\},$$
$$\omega_1^- \times \overline{\omega}_2 = \{x_{ij} = (ih_1, jh_2), i = 0, 1, \ldots, M-1, \\ j = 0, 1, \ldots, N\}.$$

Examinemos ahora el concepto de función reticular. Sea $\overline{\omega}$ una red introducida en una región unidimensional y x_i son los nodos de la red. La función $y = y(x_i)$ de argumento

discreto x_i, se llama *función reticular* definida sobre la red $\bar{\omega}$. Análogamente se define una función reticular sobre toda red $\bar{\omega}$ introducida en la región de cambio del argumento continuo. Por ejemplo, si x_{ij} es un nodo de la red $\bar{\omega}$ en un dominio bidimensional, entonces $y = y(x_{ij})$. Es evidente que las funciones reticulares pueden ser consideradas como funciones cuyo argumento son los números enteros que indican los nodos de la red. Así, se puede escribir, $y = y(x_i) = y(i)$, $y = y(x_{ij}) = y(i, j)$. Algunas veces para representar las funciones reticulares, nosotros utilizaremos la siguiente escritura: $y(x_i) = y_i$, $y(x_{ij}) = y_{ij}$.

La función reticular y_i se puede representar en forma de un vector, considerando los valores de la función como las componentes del vector $Y = (y_0, y_1, \ldots, y_N)$. En este ejemplo y_i está dada sobre la red $\bar{\omega} = \{x_i, i = 0, 1, \ldots, N\}$ que contiene $N + 1$ nodos, y el vector Y tiene dimensión $N + 1$. Si $\bar{\omega}$ es una red en un rectángulo $(\bar{\omega} = \{x_{ij} = (ih_1, jh_2), i = 0, 1, \ldots, M, j = 0, 1, \ldots, N\})$, entonces a la función reticular y_{ij}, prefijada sobre $\bar{\omega}$ le corresponde el vector $Y = (y_{00}, \ldots, y_{M0}, y_{01}, \ldots, y_{M1}, \ldots, y_{0N}, \ldots \ldots, y_{MN})$ de dimensión $(M + 1)(N + 1)$. Al mismo tiempo los nodos de la red $\bar{\omega}$ se consideran ordenadas por las filas de la red.

Nosotros hemos examinado las funciones reticulares escalares, o sea, aquellas funciones cuyos valores en cada nodo de la red son números. Citemos ahora ejemplos de *funciones reticulares vectoriales* cuyos valores en los nodos son vectores. Si en el ejemplo examinado más arriba, denotamos mediante $Y(x_2(j)) = Y_j$ el vector cuyos componentes son los valores de una función reticular y_{ij} en los nodos $x_{0j}, x_{1j}, \ldots, x_{Mj}$, de la j-ésima fila de la red $\bar{\omega}$: $Y_j = (y_{0j}, y_{1j}, \ldots, y_{Mj})$, $j = 0, 1, \ldots, N$, entonces obtenemos una función reticular vectorial Y_j definida sobre la red $\bar{\omega}_2 = \{x_2(j) = jh_2, j = 0, 1, \ldots, N\}$.

Si la función, prefijada sobre la red, toma valores complejos, entonces dicha función reticular se llama *compleja*.

2. Derivadas de diferencias y algunas identidades de diferencias. Sea dada una red $\bar{\omega}$. El conjunto de todas las funciones reticulares sobre $\bar{\omega}$, constituye un espacio vectorial con la suma y el producto de funciones por un escalar, definidos de manera evidente. En el espacio de las funciones reticulares se pueden definir operadores de diferencias o reticula-

res. Un operador Λ que transforma la función reticular y en la función reticular $f = \Lambda y$, se llama operador *reticular* o *de diferencias*. El conjunto de nodos de la red utilizados para escribir un operador de diferencias en un nodo de dicha red, se llama *molde* de este operador.

El operador de diferencias más sencillo es el operador de diferenciación de diferencias de una función reticular, el cual genera las derivadas de diferencias. Definamos las derivadas de diferencias.

Sea Ω una red uniforme con paso h introducida sobre la recta $-\infty < x < \infty$: $\Omega = \{x_i = a + ih,\ i = 0, \pm 1, \pm 2, \ldots\}$. Las derivadas de diferencias de primer orden para la función reticular $y_i = y(x_i)$, $x_i \in \Omega$ se definen por las fórmulas

$$\Lambda_1 y_i = y_{\bar{x},\, i} = \frac{y_i - y_{i-1}}{h}, \qquad \Lambda_2 y_i = y_{x,\, i} = \frac{y_{i+1} - y_i}{h} \tag{1}$$

y se llaman *derivadas izquierda* y *derecha* respectivamente. Se utiliza también la *derivada central*

$$\Lambda_3 y_i = y_{\overset{\circ}{x},\, i} = \frac{y_{i+1} - y_{i-1}}{2h} = 0{,}5\,(\Lambda_1 + \Lambda_2)\, y_i. \tag{2}$$

Si la red no es uniforme, entonces para las derivadas de diferencias de primer orden se aplican las siguientes notaciones:

$$y_{\bar{x},\, i} = \frac{y_i - y_{i-1}}{h_i}, \quad y_{x,\, i} = \frac{y_{i+1} - y_i}{h_{i+1}}, \quad y_{\hat{x},\, i} = \frac{y_{i+1} - y_i}{\hbar_i}$$
$$y_{\overset{\circ}{x},\, i} = 0{,}5\,(y_{\bar{x},\, i} + y_{x,\, i}), \quad \hbar_i = 0{,}5\,(h_i + h_{i+1}). \tag{3}$$

De las definiciones (1) y (3) se deducen las siguientes relaciones:

$$y_{x,\, i} = y_{\bar{x},\, i+1}, \tag{4}$$

$$y_{x,\, i} = \frac{\hbar_i}{h_{i+1}}\, y_{\hat{x},\, i}, \tag{5}$$

y también las igualdades

$$y_i = y_{i+1} - h_{i+1} y_{x,\, i} = y_{i-1} + h_i y_{\bar{x},\, i}. \tag{6}$$

Los operadores de diferencias Λ_1, Λ_2 y Λ_3 tienen moldes constituidos por dos puntos y se utilizan para aproximar la primera derivada $Lu = u'$ de una función $u = u(x)$ de una variable. Además, los operadores Λ_1 y Λ_2 aproximan el operador L sobre las funciones suaves con error $O(h)$ y Λ_3 con error $O(h^2)$.

Las derivadas de diferencias de n-ésimo orden se definen como funciones reticulares obtenidas al calcular la primera derivada de diferencias de la función ya derivada de diferencias hasta el orden $n - 1$. Citemos ejemplos de derivadas de diferencias de segundo orden:

$$y_{\bar{x}x,i} = \frac{y_{\bar{x},i+1} - y_{\bar{x},i}}{h} = \frac{1}{h^2}(y_{i-1} - 2y_i + y_{i+1}),$$

$$y_{\overset{\circ}{x}\overset{\circ}{x},i} = \frac{y_{\overset{\circ}{x},i+1} - y_{\overset{\circ}{x},i-1}}{2h} = \frac{1}{4h^2}(y_{i-2} - 2y_i + y_{i+2}),$$

$$y_{\bar{x}\hat{x},i} = \frac{1}{\hbar_i}(y_{\bar{x},i+1} - y_{\bar{x},i}) = \frac{1}{\hbar_i}(y_{x,i} - y_{\bar{x},i}) = \frac{1}{\hbar_i}\left(\frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i}\right),$$

las cuales se utilizan para aproximar la segunda derivada $Lu = u''$ de la función $u = u(x)$. En el caso de una red uniforme el error de la aproximación es igual a $O(h^2)$. Los operadores de diferencias correspondientes poseen un molde tripuntual. Para aproximar la cuarta derivada $Lu = u^{IV}$ se utiliza la derivada de diferencias de cuarto orden $y_{\bar{x}x\bar{x}x,i} = \frac{1}{h^4}(y_{i-2} - 4y_{i-1} + 6y_i - 4y_{i+1} + y_{i+2})$. Análogamente, para aproximar derivadas de n-ésimo orden se utilizan derivadas de diferencias de orden n.

No presenta dificultad definir las derivadas de diferencias de funciones reticulares de varias variables.

Para transformar las expresiones que contengan derivadas de diferencias de funciones reticulares, necesitaremos fórmulas de diferenciación de diferencias para el producto de funciones reticulares y fórmulas de sumación por partes. Estas fórmulas son el análogo de las correspondientes fórmulas del cálculo diferencial.

1) *Fórmulas de diferenciación de diferencias del producto.* Utilizando la definición (3) de derivadas de diferencias, no es difícil verificar, que tienen lugar las identidades:

$$(uv)_{\bar{x},i} = u_{\bar{x},i}v_{i-1} + u_i v_{\bar{x},i} = u_{\bar{x},i}v_i + u_{i-1}v_{\bar{x},i} = u_{\bar{x},i}v_i + u_i v_{\bar{x},i} - h_i u_{\bar{x},i}v_{\bar{x},i},$$

$$(uv)_{x,i} = u_{x,i}v_{i+1} + u_i v_{x,i} = u_{x,i}v_i + u_{i+1}v_{x,i} = u_{x,i}v_i + u_i v_{x,i} + h_{i+1}u_{x,i}v_{x,i},$$

$$(uv)_{\hat{x},i} = u_{\hat{x},i}v_{i+1} + u_i v_{\hat{x},i} = u_{\hat{x},i}v_i + u_{i+1}v_{\hat{x},i} = u_{\hat{x},i}v_i + u_i v_{\hat{x},i} + \hbar_i u_{\hat{x},i}v_{\hat{x},i}.$$

Utilizando (4) y (5) se puede escribir la última identidad en la forma:

$$(uv)_{\hat{x},i} = u_{\hat{x},i} v_i + \frac{h_{i+1}}{\hbar_i} u_{i+1} v_{\bar{x},i+1} \qquad (7)$$

2) *Fórmulas de sumación por partes.* Multiplicando (7) por $\hbar_i$ y sumando la relación obtenida con respecto a i desde $m+1$ hasta $n-1$, encontramos que

$$\sum_{i=m+1}^{n-1} (uv)_{\hat{x},i} \hbar_i = u_n v_n - u_{m+1} v_{m+1} =$$

$$= \sum_{i=m+1}^{n-1} u_{\hat{x},i} v_i \hbar_i + \sum_{i=m+1}^{n-1} u_{i+1} v_{\bar{x},i+1} h_{i+1}.$$

Utilizando (6) obtenemos la relación $v_{m+1} = v_m + h_{m+1} v_{x,m} = v_m + h_{m+1} v_{\bar{x},m+1}$, la cual sustituye en la igualdad hallada. Como resultado tendremos

$$u_n v_n - u_{m+1} v_m = \sum_{i=m+1}^{n-1} u_{\hat{x},i} v_i \hbar_i + \sum_{i=m}^{n-1} u_{i+1} v_{\bar{x},i+1} h_{i+1}.$$

El cambio del índice de sumación $i' = i - 1$ en la segunda suma del miembro segundo, da la siguiente fórmula de sumación por miembros:

$$\sum_{i=m+1}^{n-1} u_{\hat{x},i} v_i \hbar_i = - \sum_{i=m+1}^{n} u_i v_{\bar{x},i} h_i + u_n v_n - v_{m+1} v_m. \qquad (8)$$

Utilizando (6) es fácil obtener de (8) una fórmula de sumación más por miembros

$$\sum_{i=m+1}^{n-1} u_{\bar{x},i} v_i h_i = - \sum_{i=m}^{n-1} u_i v_{\hat{x},i} \hbar_i + u_{n-1} v_n - u_m v_m. \qquad (9)$$

De la fórmula (8) se deduce que la función u_i debe estar definida para $m+1 \leqslant i \leqslant n$, y la función v_i — para $m \leqslant i \leqslant n$. Sea ahora y_i una función reticular dada, para $m \leqslant i \leqslant n$. Entonces la función $u_i = y_{\bar{x},i}$ está definida para $m+1 \leqslant i \leqslant n$. Sustituyendo u_i en (8), obtenemos la siguiente identidad:

$$\sum_{i=m+1}^{n-1} y_{\bar{x}\hat{x},i} v_i \hbar_i = - \sum_{i=m+1}^{n} y_{\bar{x},i} v_{\bar{x},i} h_i + y_{\bar{x},n} v_n - y_{x,m} v_m. \qquad (10)$$

Tiene lugar el siguiente lema:

LEMA 1. *Sea $\bar{\omega} = \{x_i, i = 0, 1, \ldots, N, x_0 = 0, x_N = l\}$, una red no uniforme arbitraria y sea y_i una función reticular dada sobre $\bar{\omega}$, la cual se anula para $i = 0$, $i = N$. Para esta función se cumple la igualdad*

$$\sum_{i=1}^{N-1} y_{\bar{x}\hat{x}, i} y_i \hbar_i = - \sum_{i=1}^{N} (y_{\bar{x}, i})^2 h_i.$$

La afirmación del lema 1 se deduce de forma evidente de la identidad (10).

COROLARIO. *Si $\bar{\omega}$ es una red uniforme,* $y_0 = y_N = 0$ y $y_i \not\equiv 0$, entonces

$$\sum_{i=1}^{N-1} y_{\bar{x}x, i} y_i h = - \sum_{i=1}^{N} y_{\bar{x}, i}^2 h < 0.$$

Con esto terminamos el examen de las fórmulas de diferencias. Algunas otras fórmulas serán examinadas en el cap. V.

Las identidades obtenidas se utilizan no solamente para la transformación de las expresiones de diferencias. Ellas se aplican frecuentemente, por ejemplo, para cálculos de distinto tipo de sumas finitas y series.

Citemos un ejemplo. Se exige calcular la suma $S_n = \sum_{i=1}^{n-1} ia^i$, $a \neq 1$. Introduzcamos las siguientes funciones reticulares prefijadas sobre la red uniforme $\bar{\omega} = \{x_i = i, i = 0, 1, \ldots, N, h = 1\}$:

$$v_i = i, \quad u_i = (a^i - a^n)/(a - 1). \qquad (11)$$

Sobre la red indicada la fórmula de sumación por miembros (8) para cualesquiera funciones reticulares, tiene la forma ($m = 0$)

$$\sum_{i=1}^{n-1} u_{x, i} v_i = - \sum_{i=1}^{n} u_i v_{\bar{x}, i} + u_n v_n - u_1 v_0.$$

Teniendo en cuenta que para funciones (11) son ciertas las relaciones $v_0 = u_n = 0$, $v_{\bar{x}, i} = 1$, $u_{x, i} = a^i$, de aquí obtenemos $S_n = \sum_{i=1}^{n-1} ia^i = - \sum_{i=1}^{n} \frac{a^i - a^n}{a - 1} = \frac{a^n (n (a-1) - a) + a}{(a-1)^2}$.

La suma buscada ha sido hallada.

3. Ecuaciones reticulares y en diferencias. Sea $y_i = y(i)$ una función reticular del argumento discreto i. A su vez, los valores de la función reticular $y(i)$ forman un conjunto discreto. Sobre este conjunto se puede definir una función reticular que al ser igualada a cero obtenemos una ecuación con respecto a la función $y(i)$ — *ecuación reticular*. Un caso especial de ecuación reticular es una *ecuación en* o *de diferencias*. Precisamente las ecuaciones de diferencias serán el objeto fundamental de investigación en nuestro libro.

Las ecuaciones reticulares se obtienen al aproximar ecuaciones diferenciales e integrales sobre una red.

Citemos primeramente ejemplos de aproximaciones de diferencias de ecuaciones diferenciales ordinarias.

Así, las ecuaciones diferenciales de primer orden $\frac{du}{dx} = f(x)$, $x > 0$, se cambian por ecuaciones en diferencias de primer orden $\frac{y_{i+1} - y_i}{h} = f(x_i)$, $x_i = ih$, $i = 0, 1, \ldots$ o $y_{i+1} = y_i + hf(x_i)$, donde h es el paso de la red $\omega = \{x_i + ih, \ i = 0, 1, \ldots\}$. La función buscada es la función reticular $y_i - y(i)$.

En la aproximación de diferencias de una ecuación de segundo orden $\frac{d^2u}{dx^2} = f(x)$, obtenemos la ecuación en diferencias de segundo orden $y_{i+1} - 2y_i + y_{i-1} = \varphi_i$, $\varphi_i = h^2 f_i$, $f_i = f(x_i)$, $x_i = ih$. Si aproximamos la ecuación de tipo general $(ku')' + ru' - qu = f(x)$, en un molde tripuntual (x_{i-1}, x_i, x_{i+1}), entonces obtendremos una ecuación en diferencias de segundo orden con coeficientes variables del tipo $a_i y_{i-1} - c_i y_i + b_i y_{i+1} = -\varphi_i$, $i - 0, 1, \ldots$, donde a_i, c_i, b_i y φ_i son las funciones reticulares dadas, e y_i es la función reticular buscada.

La aproximación sobre un retículo de la ecuación de cuarto orden $(ku'')'' = f(x)$ conduce a una ecuación en diferencias de cuarto orden y tiene la forma

$$a_i^{(2)} y_{i-2} + a_i^{(1)} y_{i-1} + c_i y_i + b_i^{(1)} y_{i+1} + b_i^{(2)} y_{i+2} = \varphi_i.$$

Para la aproximación de diferencias de las derivadas u', u'' y u''' se pueden utilizar moldes con un número grande de nodos. Esto conduce a las ecuaciones en diferencias de orden más alto.

La ecuación lineal respecto a la función reticular $y(i)$ (función del argumento entero i)

$$a_0(i)\, y(i) + a_1(i)\, y(i+1) + \ldots + a_m(i)\, y(i+m) = f(i), \qquad (12)$$

donde $a_0(i) \neq 0$, $a_m(i) \neq 0$, y $f(i)$ es una función reticular dada que se llama *ecuación en diferencias de m-ésimo orden.*

Si (12) no contiene $y(i)$, pero contiene $y(i+1)$, entonces la sustitución de la variable independiente $i+1$ por i' reduce esta ecuación a una ecuación de orden $m-1$.

En esto consiste una de las diferencias entre las ecuaciones reticulares y las diferenciales, donde la sustitución de la variable independiente no cambia la ecuación.

Sea $F(i, y(i), y(i+1), \ldots, y(i+m))$ una función reticular no lineal. Entonces $F(i, y(i), y(i+1), \ldots$ $\ldots, y(i+m)) = 0$ *es la ecuación en o de diferencias no lineal de m-ésimo orden*, si F depende explícitamente de $y(i)$ e $y(i+m)$.

Para comodidad al comparar con las ecuaciones diferenciales, introduzcamos *las diferencias (derechas) para las funciones reticulares*: $\Delta y_i = y_{i+1} - y_i$, $\Delta^2 y_i = \Delta(\Delta y_i)$, $\ldots$ $\ldots$, $\Delta^{k+1} y_i = \Delta(\Delta^k y_i)$, $k = 1, 2, \ldots$.

Entonces (12) se puede escribir en la forma

$$\alpha_0(i)\, y(i) + \alpha_1(i)\, \Delta y_i + \ldots + \alpha_m(i)\, \Delta^m y_i = f_i, \quad (12')$$

donde $\alpha_m(i) = a_m(i) \neq 0$ y además, el coeficiente a_0 de y_0, también es diferente de cero.

La ecuación en diferencias (12′) es el análogo formal de la ecuación diferencial de orden m:

$$\alpha_0 u + \alpha_1 \frac{du}{\partial x} + \ldots + \alpha_{m-1} \frac{d^{m-1}}{dx^{m-1}} + \alpha_m \frac{d^m u}{dx^m} = f(x),$$

donde $\alpha_m \neq 0$, $\alpha_k = \alpha_k(x)$, $k = 0, 1, \ldots, m$. Sea dada la red $\omega = \{x_i = ih,\ i = 0, 1, \ldots\}$. Si denotemos

$$y_{x,i} = \frac{y_{i+1} - y_i}{h}, \quad y_{xx,i} = (y_x)_{x,i}, \ \ldots, \ y_x^{(h)} =$$
$$= y_{\underbrace{x, \ \ldots, \ x}_{k \text{ veces}}, \ i},$$

de manera tal que $y_x^{(k)} = (y_x^{(k-1)})_x$, $k \geqslant 1$, $y_{x,i}^{(0)} = y(i)$, entonces $y(i+k)$ se expresa mediante $y(i)$, $y_x^{(1)}$, $\ldots$, $y_x^{(k-1)}$ por ejemplo, $y(i+3) = y(i) + 3hy_{x,i} + 3h^2 y_{xx,i} + h^3 y_{xxx,i}$.

Luego la ecuación (12) se escribe en la forma $\bar{\alpha}_0 y(i) +$ $+ \bar{\alpha}_1(i)\, y_x(i) + \ldots + \bar{\alpha}_{m-1} y_x^{(m-1)}(i) + \bar{\alpha}_m y_x^{(m)}(i) = f_i$, donde $\bar{\alpha}_m = a_m \neq 0$ y $a_0 \neq 0$. Aquí la analogía con la ecuación diferencial de m-ésimo orden es evidente.

Similarmente se define la ecuación en diferencias respecto a la función reticular $y_{i_1, i_2} = y(i_1, i_2)$ de dos argumentos

discretos y en general, de cualquier número de argumentos. Por ejemplo, el esquema pentapuntual de diferencias «cruz» para la ecuación de Poisson $\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = -f(x_1, x_2)$ sobre la red $\omega = \{x_i = (i_1 h_1, i_2 h_2), i_1, i_2 = 0, 1, \ldots\}$, tiene la forma

$$\frac{y(i_1-1, i_2)-2y(i_1, i_2)+y(i_1+1, i_2)}{h_1^2} + \\ + \frac{y(i_1, i_2-1)-2y(i_1, i_2)+y(i_1, i_2+1)}{h_2^2} = -f_{i_1, i_2}$$

y representa una ecuación en diferencias de segundo orden respecto de cada uno de los argumentos discretos i_1 e i_2.

La ecuación reticular de tipo *general* se obtiene al aproximar la ecuación integral $u(x) = \int_0^1 K(x, s)\, u(s)\, ds + f(x)$, $0 \leqslant x \leqslant 1$, sobre la red $\bar{\omega} = \{x_i = ih, i = 0, 1, \ldots, N, hN = 1\}$.

Cambiemos la integral por la suma

$$\int_0^1 K(x, s)\, u(s)\, ds \approx h \sum_{j=0}^{N} \alpha_j K(x, jh)\, u(jh),$$

donde α_j es el coeficiente de la fórmula de cuadratura, y en lugar de la ecuación integral escribamos la ecuación reticular

$$y_i = \sum_{j=0}^{N} \alpha_j K(ih, jh)\, y_j + f_i, \quad i = 0, \quad 1, \ \ldots, \ N,$$

donde la sumación se realiza por todos los nodos de la red $\bar{\omega}$, y la función reticular y_i es la incógnita.

La ecuación reticular puede ser escrita en la forma

$$\sum_{j=0}^{N} c_{ij} y_j = f_i, \quad i = 0, \quad 1, \ \ldots, \ N. \qquad (13)$$

Ella contiene todos los valores $y_0, y_1, \ldots, y_N$ de la función reticular y se puede interpretarla como la ecuación en diferencias de orden N igual al número de nodos de la red menos uno.

La ecuación de diferencias (12) de m-ésimo orden es un tipo especial de ecuación reticular, cuando la matriz (c_{ij}) tiene elementos distintos de cero solamente sobre m diagonales paralelas a la diagonal principal.

En el caso general se puede entender por i no sólo un índice $i = 0, 1, \ldots$, sino también un multi-índice, es decir, un vector $i = (i_1, i_2, \ldots, i_p)$ con números enteros en las componentes $i_\alpha = 0, 1, 2, \ldots, \alpha = 1, 2, \ldots, p$, además $i \in \omega$, donde ω es la red.

La ecuación reticular lineal tiene la forma

$$\sum_{j \in \omega} c_{ij} y_j = f_i, \quad i \in \omega, \tag{14}$$

donde la sumación se realiza por todos los nodos de la red ω, f_i es la función reticular prefijada e y_i, la función reticular buscada.

Si renumeramos todos los nodos de la red, entonces se puede escribir $y_i = y(i)$, donde i es el número del nodo, $i = 0, 1, 2, \ldots, N$. Debido a esto la ecuación reticular (14) toma la forma (13).

Evidentemente que esto es un sistema de ecuaciones algebraicas lineales de orden $N + 1$ con matriz (c_{ij}). Por lo tanto todo sistema de ecuaciones algebraicas lineales se puede interpretar como una ecuación reticular e inversamente.

Si $y(i)$ es una función reticular *vectorial*, entonces se habla de *la ecuación reticular (de diferencias) vectorial de m-ésimo orden.*

Sea $F(i, y_0, y_1, \ldots, y_N)$ la función dada (en general, no lineal) de $N + 2$ argumentos $i, y_0, y_1, \ldots, y_N$. Igualándola a cero, obtenemos la ecuación reticular no lineal $F(i, y_0, y_1, \ldots, y_N) = 0$, $i = 0, 1, \ldots, N$, cuya solución es la función reticular $y(i)$ que convierte esta ecuación en identidad.

Examinemos la función reticular $\mathscr{F}(i) = F(i, y_0, y_1, \ldots, y_N)$, $i = 0, 1, \ldots, N$. De aquí se ve, que la función F prefija un cierto operador reticular que convierte la función reticular $y(i)$ en la función reticular $\mathscr{F}(i)$.

Si F es la función lineal, entonces obtenemos la ecuación (14), la cual, obviamente, se puede escribir en la forma operacional $Ay = f$, donde A es el operador lineal con matriz (c_{ij}), e y es el vector en el espacio de las funciones reticulares.

Si los coeficientes c_{ij} no dependen de j, entonces (14) se llama *ecuación reticular con coeficientes constantes.*

Aunque en este libro se le concede atención fundamental a la solución numérica de las ecuaciones de diferencias que se obtienen al aproximar en diferencias las ecuaciones diferenciales de tipo elíptico, los métodos iterativos son aplicables

para cualquier ecuación reticular lineal, es decir, para todo sistema de ecuaciones algebraicas lineales. Por eso la teoría de los métodos iterativos aquí expuesta tiene un carácter general. La particularidad de las ecuaciones reticulares consiste en que son un sistema de alto orden, y además el orden de la ecuación aumenta al condensarse la red (el número de incógnitas es igual al número N de nodos de la red, $N = O\left(\frac{1}{h^p}\right)$ en el caso p-dimensional, donde h es el paso de la red).

4. **Problema de Cauchy y problemas de contorno para ecuaciones en diferencias.** Citemos algunos ejemplos complementarios de ecuaciones en diferencias y detengámonos en el planteamiento de los problemas para las ecuaciones en diferencias.

Observemos, que los ejemplos más simples de ecuaciones en diferencias de primer orden son las fórmulas para los términos de las progresiones aritmética y geométrica:

$$y_{i+1} = y_i + d, \quad y_{i+1} = qy_i, \quad i = 0, 1, \ldots$$

La solución de una ecuación de primer orden puede ser hallada, si se da una condición inicial para $i = 0$ (problema de Cauchy).

La solución $y(i + m)$ de una ecuación de diferencias de m-ésimo orden está determinada completamente por los valores de $y(i)$, prefijados en m puntos arbitrarios, pero ubicados consecutivamente, i_0, $i_0 + 1$, $\ldots$, $i_0 + m - 1$. En efecto, ya que $a_m(i) \neq 0$, entonces de (12) encontramos $y(i + m) = b_{m-1}(i)\, y(i + m - 1) + \ldots + b_0(i)\, y(i) + \varphi(i)$. Poniendo aquí sucesivamente $i = i_0, i_0 + 1, \ldots$, hallamos los valores $y(i)$ para $i \geqslant i_0$. Análogamente, expresando de (12) $y(i)$ mediante $y(i + 1), \ldots, y(i + m)$ y poniendo sucesivamente $i = i_0 - 1, i_0 - 2, \ldots$, hallamos $y(i)$ para $i \leqslant i_0 - 1$. Si en la ecuación (12) se exige determinar $y(i)$ para $i \geqslant 0$, entonces es suficiente dar el valor en m nodos (condiciones iniciales) $y(0) = y_0$, $y(1) = y_1$, $\ldots$ $\ldots$, $y(m - 1) = y_{m-1}$.

Añadiendo estas condiciones a la ecuación (12), obtenemos el problema de Cauchy o problema con datos iniciales para la ecuación de diferencias de m-ésimo orden.

Para las ecuaciones de primer orden ($m = 1$), como hemos visto, es suficiente dar una condición inicial.

Las ecuaciones de diferencias no lineales se obtienen al resolver ecuaciones diferenciales no lineales. Examinemos,

por ejemplo, la ecuación diferencial

$$\frac{du}{dx}=f(x, u), \quad x>0, \quad u(0)=\mu_1$$

(problema de Cauchy). Sustituyendo esta ecuación por el esquema de Euler (esquema explícito), obtendremos la ecuación de diferencias de primer orden $y_{i+1} = y_i + hf(x_i, y_i)$ donde $i \geqslant 0$ y $y_0 = \mu_1$.

Si sustituyomos la derivada du/dx para $x = x_i = ih$ por la relación izquierda en diferencias, entonces obtendremos la ecuación de diferencias no lineal respecto a y_i de primer orden $y_i = y_{i-1} + hf(x_i, y_i)$, donde $i > 0$ y $y_0 = \mu_1$. Para determinar y_i es necesario resolver la ecuación no lineal $\varphi(y_i) = y_i - hf(x_i, y_i) = y_{i-1}$.

Examinemos ahora un ejemplo de ecuación de diferencias de segundo orden. Supongamos que se exige calcular las integrales $I_k(\varphi) = \int_0^\pi \frac{\cos k\psi - \cos k\varphi}{\cos\psi - \cos\varphi}\, d\psi$, $k = 0, 1, 2, \ldots$

Ante todo notemos, que $I_0(\varphi) = 0$ e $I_1(\varphi) = \pi$. Transformemos la expresión $[\cos(k+1)\psi - \cos(k+1)\varphi] + [\cos(k-1)\psi - \cos(k-1)\varphi] = 2\cos k\psi\cos\psi - 2\cos k\varphi \times \cos\varphi = 2(\cos k\psi - \cos k\varphi)\cos\varphi + 2(\cos\psi - \cos\varphi) \times \cos k\psi$. Aprovechando esta expresión, obtenemos:

$$I_{k+1}(\varphi)+I_{k-1}(\varphi)=2\cos\varphi I_k(\varphi)+ +2\int_0^\pi \cos k\psi\, d\psi=2\cos\varphi I_k(\varphi), \quad k\geqslant 1.$$

De esta manera, el cálculo de las integrales $I_k(\varphi)$ se reduce a la solución del problema de Cauchy para la ecuación de diferencias de segundo orden

$$I_{k+1}(\varphi) - 2\cos\varphi I_k(\varphi) + I_{k-1}(\varphi) = 0,$$
$$k \geqslant 1, \quad I_0(\varphi) = 0, \quad I_1(\varphi) = \pi. \tag{15}$$

Examinemos otro ejemplo más. Se exige hallar la solución de un problema de contorno para el sistema de ecuaciones diferenciales ordinarias de primer orden

$$\frac{du}{dx}=Au+f(x), \quad 0<x<l, \tag{16}$$

con $Bu = \mu_1$ si $x = 0$ y $Cu = \mu_2$ para $x = l$. Aquí $u(x) = (u_1(x), u_2(x), \ldots, u_M(x))$ es la función vectorial de dimensión M, $A = A(x)$ es una matriz cuadrada de dimen-

sión $M \times M$, B y C son las matrices rectangulares de dimensión $M_1 \times M$ y $M_2 \times M$ respectivamente, donde $M_1 + M_2 = M$. Los vectores $f(x)$, μ_1 y μ_2 son dados y tienen dimensiones M, M_1 y M_2 respectivamente.

Introduciendo la red uniforme $\overline{\omega} = \{x_i = ih, \ i = 0, 1, \ldots, N, \ h = l/N$ en el segmento $0 \leqslant x \leqslant l$ y definiendo sobre ella la función vectorial reticular $Y_i = (y_1(i), y_2(i), \ldots, y_M(i))$, le ponemos en correspondencia al problema (16) el esquema de diferencias más sencillo.

$$Y_{i+1} - (E + hA_i) Y_i = F_i, \quad 0 \leqslant i \leqslant N-1,$$
$$BY_0 = \mu_1, \qquad CY_N = \mu_2, \tag{17}$$

donde $F_i = hf(x_i)$. Esto es un ejemplo de ecuación vectorial lineal de diferencias de primer orden con M_1 condiciones para $i = 0$ y M_2 condiciones para $i = N$. Por lo tanto, para el sistema de ecuaciones de diferencias de primer orden, tenemos un problema de contorno.

Para las ecuaciones de segundo orden son más típicos los problemas de contorno. Examinemos, por ejemplo, el primer problema de contorno

$$\frac{d^2u}{dx^2} - q(x)\,u = -f(x), \quad 0 < x < l,$$
$$u(0) = \mu_1 \quad u(l) = \mu_2 \quad q(x) \geqslant 0. \tag{18}$$

Elijamos la red $\overline{\omega} = \{x_i = ih, \ i = 0, 1, \ldots, N, \ h = l/N\}$ y pongamos al problema (18) en correspondencia el problema de contorno en diferencias,

$$y_{\bar{x}x,\,i} - d_i y_i = -\varphi_i, \quad 0 < i < N,$$
$$y_0 = \mu_1, \qquad y_N = \mu_2 \tag{19}$$

donde $d_i = q(x_i)$ y $\varphi_i = f(x_i)$ para $q(x)$ y $f(x)$ suaves. Este problema es un caso particular del problema de contorno para la ecuación de diferencias de segundo orden

$$-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = \varphi_i,$$
$$1 \leqslant i \leqslant N-1, \quad y_0 = \mu_1, \qquad y_N = \mu_2 \tag{20}$$

para $a_i = b_i = 1/h^2$ y $c_i = d_i + 2/h^2$.

El esquema de diferencias (20) se puede escribir en la forma

$$\mathcal{A}Y = F, \tag{21}$$

donde $Y = (y_1, y_2, \ldots, y_{N-1})$ es el vector de dimensión $N-1$ incógnito, $F = \left(\varphi_1 + \frac{1}{h^2}\mu_1, \varphi_2, \ldots, \varphi_{N-2}, \varphi_{N-1} + \right.$ $\left. + \frac{1}{h^2}\mu_2\right)$ es el vector de dimensión $N-1$ conocido y $\mathcal{A}$ es la matriz cuadrada tridiagonal del tipo

$$\mathcal{A} = \begin{Vmatrix} c_1 & -b_1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ -a_2 & c_2 & -b_2 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -a_3 & c_3 & -b_3 & \ldots & 0 & 0 & 0 \\ . & . & . & . & \ldots & . & . & . \\ 0 & 0 & 0 & 0 & \ldots & c_{N-3} & -b_{N-3} & 0 \\ 0 & 0 & 0 & 0 & \ldots & -a_{N-2} & c_{N-2} & -b_{N-2} \\ 0 & 0 & 0 & 0 & \ldots & 0 & -a_{N-1} & c_{N-1} \end{Vmatrix} \qquad (22)$$

De aquí se ve que el problema de contorno para la ecuación de diferencias de segundo orden (20) representa un sistema de ecuaciones algebraicas lineales de tipo especial. Si el problema de Cauchy para la ecuación de diferencias de segundo orden es siempre soluble, entonces el primer problema de contorno (20) es soluble para todo segundo miembro derecho solamente, cuando la matriz $\mathcal{A}$ del sistema (21) no es degenerada.

Los problemas de contorno para ecuaciones en diferencias de m-ésimo orden reducen a los sistemas de ecuaciones algebraicas lineales con una matriz que no posee más de $m + 1$ elementos no nulos en cada fila.

Al aproximar ecuaciones en derivadas parciales nosotros llegamos también a un sistema de ecuaciones en diferencias o sencillamente algebraicas con una matriz especial. Ya que el número de incógnitas en tal sistema frecuentemente es igual al número de nodos de la red, entonces en la práctica nos encontramos con sistemas de orden muy alto (con decenas y hasta centenas de miles de incógnitas). Otras singulares de estos sistemas son el enrarecimiento de la matriz y la estructura en forma de banda, es decir, la distribución especial de los elementos no nulos. Estas singularidades, por una parte, facilitan la solución de los problemas indicados y, por otra parte, exigen la creación de métodos especiales de solución que tengan en cuenta las particularidades del problema. Por eso, no debe asombrarnos que los métodos clásicos del álgebra lineal, resulten a menudo no efectivos

para resolver ecuaciones de diferencias y que no exista un método universal que permita resolver eficientemente cualquier ecuación de diferencias.

En la actualidad se utilizan dos tipos de métodos de solución de sistemas de ecuaciones algebraicas lineales: 1) métodos directos; 2) métodos iterativos o métodos de aproximaciones sucesivas. Como regla general, los métodos directos están orientados hacia la solución de una clase bastante estrecha de ecuaciones reticulares, pero ellos permiten encontrar la solución con un gasto muy pequeño de trabajo computacional. Los métodos iterativos permiten resolver ecuaciones más complejas y con frecuencia, en calidad de etapa fundamental del algoritmo, contienen métodos directos de solución de ecuaciones de diferencias especiales. El hecho de que las ecuaciones de diferencias están mal acondicionadas, conduce a la necesidad de elaborar procesos iterativos rápidamente convergentes y a la distinción del dominio de efectividad de cada método.

En una serie de casos, por ejemplo para una ecuación lineal con coeficientes constantes con respecto a una función reticular de un argumento, la solución puede ser hallada en forma cerrada. Tales métodos de solución de ecuaciones reticulares serán examinados en el § 3 de este capítulo.

§ 2. Teoría general de las ecuaciones lineales en diferencias

1. Propiedades de las soluciones de una ecuación homogénea. En este párrafo será examinada la teoría general de ecuaciones en diferencias lineales de m-ésimo orden con coeficientes variables,

$$a_m(i)\, y(i+m) + \ldots + a_0(i)\, y(i) = f_i,$$

donde $a_m(i)$ y $a_0(i)$ son distintas de cero para cualquier i. Ocupémonos primeramente de investigar la ecuación homogénea

$$a_m(i)\, y(i+m) + \ldots + a_0(i)\, y(i) = \sum_{k=0}^{m} a_k(i)\, y(i+k) = 0. \quad (1)$$

Consideraremos que los coeficientes $a_k(i)$, $k = 0, 1, \ldots, m$, tienen valores finitos para todos los valores de i examinados.

Cada solución particular de la ecuación (1) se determina por los valores de la función $y(i)$ en m puntos arbitrarios, pero situados consecutivamente, $i_0, i_0+1, \ldots, i_0+m-1$.

TEOREMA 1. *Si $v_1(i), v_2(i), \ldots, v_p(i)$ son soluciones de la ecuación (1), entonces la función*

$$y(i) = c_1 v_1(i) + c_2 v_2(i) + \ldots + c_p v_p(i), \tag{2}$$

donde $c_1, c_2, \ldots, c_p$ son constantes arbitrarias, es también solución de la ecuación (1).

En efecto, en virtud de la condición del teorema, tienen lugar las igualdades

$$\sum_{k=0}^{m} a_k(i) v_l(i+k) = 0 \quad l = 1, 2, \ldots, p. \tag{3}$$

Sustituyamos (2) en (1):

$$\sum_{k=0}^{m} a_k(i) y(i+k) = \sum_{k=0}^{m} a_k(i) \sum_{l=1}^{p} c_l v_l(i+k)$$

y cambiemos el orden de sumación en el segundo miembro de la igualdad. Utilizando (3), obtenemos

$$\sum_{k=0}^{m} a_k(i) y(i+k) = \sum_{l=1}^{p} c_l \sum_{k=0}^{m} a_k(i) v_l(i+k) = 0$$

y, por consiguiente, la función $y(i)$, definida por (2), también es solución de la ecuación (1). El teorema está demostrado.

Introduzcamos la notación $\Delta_i(v_1, \ldots, v_p)$ para el determinante

$$\Delta_i(v_1, v_2, \ldots, v_p) = \begin{vmatrix} v_1(i) & v_1(i+1) & \ldots & v_1(i+p-1) \\ v_2(i) & v_2(i+1) & \ldots & v_2(i+p-1) \\ \cdot & \cdot & \cdot & \cdot \\ v_p(i) & v_p(i+1) & \ldots & v_p(i+p-1) \end{vmatrix}.$$

Tiene lugar el siguiente lema:

LEMA 2. *Sean $v_1(i), v_2(i), \ldots, v_m(i)$ las soluciones de la ecuación (1). El determinante $\Delta_i(v_1, \ldots, v_m)$ es idénticamente igual a cero según i, o distinto de cero para todos los valores admisibles de i.*

Realmente, ya que $v_i(i), \ldots, v_m(i)$, son soluciones de la ecuación (1), entonces son válidas las igualdades siguien-

tes:

$$a_0(i)v_1(i)+a_1(i)v_1(i+1)+\dots$$
$$\dots+a_{m-1}(i)v_1(i+m-1)=-a_m(i)v_1(i+m),$$
$$a_0(i)v_2(i)+a_1(i)v_2(i+1)+\dots$$
$$\dots+a_{m-1}(i)v_2(i+m-1)=-a_m(i)v_2(i+m),$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$a_0(i)v_m(i)+a_1(i)v_m(i+1)+\dots$$
$$\dots+a_{m-1}(i)v_m(i+m-1)=-a_m(i)v_m(i+m).$$

Resolviendo este sistema por la regla de Cramer, con respecto a $a_0(i)$ para i fijo, obtenemos

$$a_0(i)\Delta_i(v_1,\dots,v_m)=$$
$$=-a_m(i)\begin{vmatrix} v_1(i+m) & v_1(i+1) & \dots & v_1(i+m-1) \\ v_2(i+m) & v_2(i+1) & \dots & v_2(i+m-1) \\ \dots & \dots & \dots & \dots \\ v_m(i+m) & v_m(i+1) & \dots & v_m(i+m-1) \end{vmatrix}.$$

Después del respectivo reordenamiento de las columnas del determinante de la parte derecha de la igualdad obtenida, tendremos la relación $a_0(i)\Delta_i(v_1,\dots,v_m)=$ $=(-1)^m a_m(i)\Delta_{i+1}(v_1,\dots,v_m)$. Ya que $a_0(i)$ y $a_m(i)$ son distintos de cero para los valores admisibles de i, entonces de aquí se deduce la afirmación del lema.

Introduzcamos ahora el concepto de soluciones linealmente independientes de la ecuación (1). Las funciones reticulares $v_1(i)$, $v_2(i)$, $\dots$, $v_m(i)$ se llaman *soluciones linealmente independientes de la ecuación* (1), si: 1) ellas toman valores finitos y satisfacen la ecuación (1); 2) las relaciones

$$c_1v_1(i)+c_2v_2(i)+\dots+c_mv_m(i)=0 \qquad (4)$$

para constantes cualesquiera $c_1, c_2, \dots, c_m$, simultáneamente no iguales a cero, no se cumplen al menos para un i.

Para las soluciones linealmente independientes es válido el siguiente lema:

LEMA 3. *Si $v_1(i)$, $v_2(i)$, $\dots$, $v_m(i)$ son las soluciones linealmente independientes de la ecuación (1), el determinante $\Delta_i(v_1,\dots,v_m)$ es distinto de cero para todos los valores admisibles de i. Inversamente, si para las soluciones $v_1(i)$, $\dots$, $v_m(i)$ de la ecuación (1) el determinante $\Delta_i(v_1,\dots,v_m)$ es distinto de cero al menos para un valor de i, entonces*

$v_1(i), \ldots, v_m(i)$ son las soluciones linealmente independientes de la ecuación (1).

En virtud del lema 2 el determinante $\Delta_i(v_1, \ldots, v_m)$ es idénticamente igual a cero, o distinto de cero para todos los i. Sean $v_1(i), \ldots, v_m(i)$ las soluciones linealmente independientes de la ecuación (1) y supongamos que $\Delta_i(v_1, \ldots, v_m) \equiv 0$. Examinemos el sistema de ecuaciones algebraicas

$$\begin{aligned} c_1v_1(i_0) + c_2v_2(i_0) + \ldots + c_mv_m(i_0) &= 0. \\ c_1v_1(i_0+1) + c_2v_2(i_0+1) + \ldots + c_mv_m(i_0+1) &= 0. \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \qquad (5) \\ c_1v_1(i_0+m-1) + c_2v_2(i_0+m-1) + \ldots \\ \ldots + c_mv_m(i_0+m-1) &= 0. \end{aligned}$$

Ya que, por suposición, el determinante de este sistema $\Delta_{i_0}(v_1, \ldots, v_m)$ es igual a cero, entonces existe una solución $c_1, c_2, \ldots, c_m$ de este sistema, distinta de cero. Por consiguiente para los $c_1, c_2, \ldots, c_m$ hallados tienen lugar las igualdades (4) para $i = i_0, i_0+1, \ldots, i_0+m-1$. Mostremos ahora, que (4) tendrá lugar también para $i = i_0+m$. Para esto tomemos la ecuación (1) para $l = 1, 2, \ldots, m$

$$\sum_{k=0}^{m} a_k(i_0)\, v_l(i_0+k) = 0,$$

multipliquémosla por c_l y sumemos las igualdades obtenidas para $l = 1, 2, \ldots, m$. Teniendo en cuenta las igualdades (5), obtendremos

$$0 = a_m(i_0)\sum_{l=1}^{m} c_lv_l(i_0+m) + \sum_{k=0}^{m-1} a_k(i_0)\sum_{l=1}^{m} c_lv_l(i_0+k) =$$
$$= a_m(i_0)\sum_{l=1}^{m} c_lv_l(i_0+m).$$

De tal modo, está demostrada la validez de la igualdad (4) para $i = i_0+m$. Siguiendo más adelante de la misma forma, obtendremos que se cumple la relación (4) para los $c_1, c_2, \ldots, c_m$ encontrados más arriba y para todos los $i \geqslant i_0$ admisibles. Análogamente se demuestra la justeza de (4) para $i \leqslant i_0$. Por consiguiente, (4) se cumple para todos los i con $c_1, c_2, \ldots, c_m$ no nulos, lo cual contradice la independencia lineal de $v_1(i), \ldots, v_m(i)$. Por eso la suposición de que

el determinante $\Delta_i(v_1, \ldots, v_m)$ es idénticamente igual a cero según i, no es cierta.

Demostremos ahora la segunda parte del lema 3. Sea el determinante $\Delta_i(v_1, \ldots, v_m)$ distinto de cero para algún $i = i_0$. Entonces supongamos, que $v_1(i), v_2(i), \ldots, v_m(i)$ es un sistema de soluciones linealmente independientes de la ecuación (1). Esto significa que se hallan constantes $c_1, c_2, \ldots, c_m$ no iguales a cero simultáneamente, tales que la relación (4) es una identidad según i. Entonces escribamos (4) para $i = i_0, i_0 + 1, \ldots, i_0 + m - 1$ en forma del sistema (5) donde en virtud de la suposición del lema, el determinante $\Delta_{i_0}(v_1, \ldots, v_m)$ de este sistema es distinto de cero. Por eso todos los $c_1, c_2, \ldots, c_m$ deben ser iguales a cero. Así llegamos a una contradicción. El lema está demostrado.

2. Teoremas sobre las soluciones de una ecuación lineal. Primeramente demostraremos un teorema sobre la solución general de la ecuación lineal homogénea (1).

TEOREMA 2. *Si $v_1(i), v_2(i), \ldots, v_m(i)$ son las soluciones linealmente independientes de la ecuación (1), entonces la solución general de esta ecuación tiene la forma*

$$y(i) = c_1 v_1(i) + c_2 v_2(i) + \ldots + c_m v_m(i), \tag{6}$$

donde $c_1, c_2, \ldots, c_m$ son las constantes arbitrarias.

Realmente, en virtud del teorema 1, la función $y(i)$ definida por la formula (6), es la solución de la ecuación (1). Mostremos ahora que todas las soluciones de la ecuación (1) están contenidas en el conjunto de las funciones $y(i)$. En efecto, sea $u(i)$ una solución arbitraria de la ecuación (1). Ella se determina completamente prefijando sus valores iniciales en m puntos: $u(i_0), u(i_0 + 1), \ldots, u(i_0 + m - 1)$. Del conjunto de las funciones del tipo (6) elijamos aquella que posee los mismos valores. Para esto es suficiente hallar constantes $c_1, c_2, \ldots, c_m$, tales que se cumplan las m igualdades.

$$c_1 v_1(i_0) + c_2 v_2(i_0) + \ldots + c_m v_m(i_0) = u(i_0),$$
$$c_1 v_1(i_0 + 1) + c_2 v_2(i_0 + 1) + \ldots + c_m v_m(i_0 + 1) = u(i_0 + 1),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$c_1 v_1(i_0 + m - 1) + c_2 v_2(i_0 + m - 1) + \ldots + c_m v_m(i_0 + m - 1) = u(i_0 + m - 1).$$

Ya que $v_1(i), v_2(i), \ldots, v_m(i)$ son las soluciones linealmente independientes de (1), entonces en virtud del lema 3, el determinante $\Delta_{i_0}(v_1, \ldots, v_m)$ de este sistema es distinto de cero. Resolviendo este sistema con respecto a $c_1, c_2, \ldots, c_m$, obtendremos la función $y(i)$ que tiene los mismos valores iniciales que $u(i)$. Pero como los valores iniciales determinan unívocamente la solución de la ecuación (1), entonces $y(i) \equiv u(i)$. El teorema está demostrado.

Examinemos ahora la solución de la ecuación no homogénea

$$a_m(i)\, y(i+m) + \ldots + a_0(i)\, y(i) = f(i) \tag{7}$$

Tiene lugar el teorema siguiente:

TEOREMA 3. *La solución general de la ecuación (7) se representa en forma de la suma de una solución particular de dicha ecuación con la solución general de la ecuación lineal homogénea (1).*

En efecto, mostremos que cualquier solución de la ecuación (7) puede ser representada en la forma

$$y(i) = \bar{y}(i) + \bar{\bar{y}}(i), \tag{8}$$

donde $\bar{y}(i)$ es cierta solución de la ecuación (7), e $\bar{\bar{y}}(i)$ es la solución general de la ecuación homogénea (1). Sea

$$a_m(i)\, \bar{y}(i+m) + \ldots + a_0(i)\, \bar{y}(i) = f(i). \tag{9}$$

Sustituyendo (8) en (7) y teniendo en cuenta (9), tendremos para $\bar{\bar{y}}(i)$ la ecuación $a_m(i)\bar{\bar{y}}(i+m) + \ldots + a_0\bar{\bar{y}}(i) = 0$. Por consiguiente, $\bar{\bar{y}}(i)$ es la solución de la ecuación homogénea (1). El teorema está demostrado.

COROLARIO 1. *De los teoremas 2 y 3 se deduce que la solución general de la ecuación no homogénea (7) tiene la forma*

$$y(i) = \bar{y}(i) + c_1 v_1(i) + \ldots + c_m v_m(i), \tag{10}$$

donde $\bar{y}(i)$ es una solución particular de la ecuación (7), y $v_1(i), v_2(i), \ldots, v_m(i)$ son las soluciones linealmente independientes de la ecuación homogénea (1), siendo $c_1, \ldots, c_m$ constantes arbitrarias.

COROLARIO 2. Utilizando el lema 3, se le puede dar otra formulación al corolario 1: *la solución de la ecuación (7) tiene la forma (10), donde las soluciones particulares $v_1(i), \ldots, v_m(i)$ de la ecuación homogénea, son tales que $\Delta_i(v_1, \ldots, v_m) \neq 0$ al menos para un valor de i.*

COROLARIO 3. *Si el miembro derecho $f(i)$ de la ecuación (7) es la suma de dos funciones $f(i) = f^{(1)}(i) + f^{(2)}(i)$, entonces una solución particular de la ecuación (7) se puede representar en la forma $\bar{y}(i) = \bar{y}^{(1)}(i) + y^{(2)}(i)$, donde $\bar{y}^{-\alpha}(i)$ es una solución particular de la ecuación (7) con miembro derecho $f^{(\alpha)}(i)$, $\alpha = 1, 2$.*

3. Método de variación de las constantes. Los teoremas demostrados más arriba dan la estructura de la solución general de la ecuación lineal no homogénea de diferencias (7). Examinemos ahora las siguientes preguntas: 1) ¿cómo construir soluciones linealmente independientes de la ecuación homogénea ; 2) ¿cómo hallar la solución particular de la ecuación no homogénea?; 3) ¿de qué manera, utilizando la solución general de la ecuación no homogénea, encontrar una única solución de la ecuación (7) que satisfaga condiciones adicionales?

Estudiemos primeramente un método posible de construcción de soluciones linealmente independientes de la ecuación homogénea. Ya que la solución particular de una ecuación lineal de m-ésimo orden se determina completamente dando los valores iniciales en m puntos, por ejemplo, $i = i_0, i_0 + 1, \ldots, i_0 + m - 1$, entonces en virtud del lema 3 las soluciones buscadas de la ecuación (1) se pueden construir de la siguiente forma. Sea A una matriz no degenerada

$$A = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \ldots & a_{mm} \end{Vmatrix}.$$

Construyamos m soluciones $v_1(i), v_2(i), \ldots, v_m(i)$ de la ecuación (1), determinadas por los valores iniciales

$$v_l(i_0 + k - 1) = a_{lk}, \quad l, k = 1, 2, \ldots, m. \tag{11}$$

Entonces $\Delta_{i_0}(v_1, \ldots, v_m) = \det A \neq 0$. Por consiguiente, el problema de construir las funciones buscadas $v_1(i), \ldots$ $\ldots, v_m(i)$, está resuelto.

Vamos a examinar ahora la pregunta de como separar una solución única de la familia de soluciones (10). De (10) se deduce, que para ello hay que prefijar exactamente m condiciones para la función $y(i)$, de las cuales se determinan las constantes $c_1, c_2, \ldots, c_m$.

En el caso del problema de Cauchy, es decir, cuando están prefijadas las condiciones iniciales $y(i_0) = b_1$, $y(i_0 + 1) = b_2, \ldots, y(i + m - 1) = b_m$, la determinación de las constantes $c_1, c_2, \ldots, c_m$ se realiza sencillamente. De (10) tenemos el sistema de ecuaciones algebraicas lineales respecto a $c_1, c_2, \ldots, c_m$:

$$\begin{gathered} v_1(i_0)c_1 + v_2(i_0)c_2 + \ldots + v_m(i_0)c_m = b_1 - \overline{y}(i_0), \\ v_1(i_0 + 1)c_1 + v_2(i_0 + 1)c_2 + \ldots \\ \ldots + v_m(i_0 + 1)c_m = b_2 - \overline{y}(i_0 + 1), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ v_1(i_0 + m - 1)c_1 + v_2(i_0 + m - 1)c_2 + \ldots \\ \ldots + v_m(i_0 + m - 1)c_m = b_m - \overline{y}(i_0 + m - 1). \end{gathered} \tag{12}$$

Ya que el determinante $\Delta_{i_0}(v_1, \ldots, v_m)$ de este sistema es distinto de cero, entonces este sistema tiene la solución única $c_1, c_2, \ldots, c_m$, la cual determina completamente la solución única de la ecuación no homogénea (7).

En el caso de un problema de contorno, cuando las m condiciones adicionales para $y(i)$ están prefijadas en puntos situados no consecutivamente, nosotros llegaremos de nuevo a un sistema de ecuaciones algebraicas lineales con respecto a $c_1, c_2, \ldots, c_m$. Pero en este caso la solución de dicho sistema existirá solamente para suposiciones adicionales respecto de los coeficientes de la ecuación en diferencias.

Examinemos ahora la pregunta sobre la solución de las ecuaciones (12). Puesto que en virtud de (11), A^T es la matriz del sistema (12), entonces, eligiendo en calidad de A la matriz unidad, obtendremos la solución del sistema (12) en forma explícita:

$$c_l = b_l - \overline{y}(i_0 + l - 1), \quad l = 1, 2, \ldots, m.$$

Evidentemente, entre las soluciones particulares de la ecuación (7) no homogénea es racional elegir aquella, para la cual $\overline{y}(i_0) = \overline{y}(i_0 + 1) = \ldots = \overline{y}(i_0 + m - 1) = 0$. Entonces tendremos $c_l = b_l$, $l = 1, 2, \ldots, m$.

A tal elección de la matriz A corresponden los siguientes valores iniciales para $v_1(i), \ldots, v_m(i)$:

$$v_l(i_0 + l - 1) = 1, \quad v_l(i_0 + k - 1) = 0,$$
$$k = 1, 2, \ldots, m, \quad k \neq l, \quad l = 1, 2, \ldots, m.$$

Ocupémonos ahora de la búsqueda de soluciones particulares de la ecuación no homogénea, si son conocidas m soluciones linealmente independientes de la ecuación homogénea. Expongamos el método de encontrar una solución particular mediante la *variación de constantes* en la solución general de la ecuación homogénea.

Antes fue mostrado, que la solución general de la ecuación homogénea (1) tiene la forma $\bar{\bar{y}}(i) = c_1 v_1(i) + \ldots + c_m v_m(i)$, donde $v_1(i), \ldots, v_m(i)$ son las soluciones linealmente independientes de la ecuación (1), y $c_1, c_2, \ldots, c_m$ son las constantes arbitrarias. Ahora vamos a considerar $c_1, c_2, \ldots, c_m$ como funciones de i y pondremos el problema de elegirlas de manera tal, que la función

$$\bar{y}(i) = c_1(i) v_1(i) + \ldots + c_m(i) v_m(i) \tag{13}$$

resulte una solución particular de la ecuación no homogénea (7). Observemos, que cada función $c_l(i)$ se determina con exactitud hasta una constante, ya que $v_l(i)$ es la solución de la ecuación homogénea:

$$a_m(i) v_l(i+m) + \ldots + a_0(i) v_l(i) = 0,$$
$$l = 1, 2, \ldots, m. \tag{14}$$

Introduzcamos la siguiente notación:

$$d_k(i) = \sum_{l=1}^{m} [c_l(i+k) - c_l(i)] v_l(i+k),$$
$$k = 0, 1, \ldots, m.$$

Sustituyendo (13) en (7), efectuando las transformaciones idénticas en la expresión obtenida y teniendo en cuenta (14), tendremos

$$f(i) = \sum_{k=0}^{m} a_k(i) \bar{y}(i+k) = \sum_{k=0}^{m} a_k(i) \sum_{l=1}^{m} c_l(i+k) v_l(i+k) =$$
$$= \sum_{k=0}^{m} a_k(i) d_k(i) + \sum_{k=0}^{m} a_k(i) \sum_{l=1}^{m} c_l(i) v_l(i+k) =$$
$$= \sum_{k=0}^{m} a_k(i) d_k(i) + \sum_{l=1}^{m} c_l(i) \Big[\sum_{k=0}^{m} a_k(i) v_l(i+k) \Big] =$$
$$= \sum_{k=0}^{m} a_k(i) d_k(i) = \sum_{k=1}^{m} a_k(i) d_k(i),$$

ya que $d_0(i) \equiv 0$. La relación obtenida se cumplirá para todos los i, si aceptemos

$$d_k(i) = 0,\ k = 1, 2, \ldots, m-1,\ d_m(i) = f(i)/a_m(i). \quad (15)$$

De esta forma, el problema de construir las funciones $c_1(i), c_2(i), \ldots, c_m(i)$ se reduce a su determinación a partir de las condiciones (15), las cuales deben satisfacerse idénticamente por i.

Transformemos el sistema de ecuaciones (15). Designemos $b_l(i) = c_l(i+1) - c_l(i)$, $l = 1, 2, \ldots, m$. De la definición de $d_k(i)$, obtenemos para $k = 1, 2, \ldots, m$:

$$d_k(i) = d_{k-1}(i+1) = \sum_{l=1}^{m} [c_l(i+k) - c_l(i)]\, v_l(i+k) -$$
$$- \sum_{l=1}^{m} [c_l(i+k) - c_l(i+1)]\, v_l(i+k) = \sum_{l=1}^{m} b_l(i)\, v_l(i+k).$$

Sustituyendo aquí (15) y teniendo en cuenta la igualdad $d_0(i) = 0$, obtendremos un sistema de ecuaciones algebraicas lineales respecto a $b_l(i)$ para i fijo:

$$\begin{aligned} &b_1(i)\, v_1(i+1) + b_2(i)\, v_2(i+1) + \ldots \\ &\qquad \ldots + b_m(i)\, v_m(i+1) = 0, \\ &b_1(i)\, v_1(i+2) + b_2(i)\, v_2(i+2) + \ldots \\ &\qquad \ldots + b_m(i)\, v_m(i+2) = 0, \qquad (16) \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &b_1(i)\, v_1(i+m) + b_2(i)\, v_2(i+m) + \ldots \\ &\qquad \ldots + b_m(i)\, v_m(i+m) = \frac{f(i)}{a_m(i)}. \end{aligned}$$

El determinante del sistema (16) es igual a $\Delta_{i+1}(v_1, v_2, \ldots, v_m)$ y es distinto de cero en virtud de la independencia lineal de $v_1, v_2, \ldots, v_m$. Por eso el sistema (16) tiene la solución única:

$$b_l(i) = c_l(i+1) - c_l(i) =$$
$$= (-1)^{m+l} \frac{f(i)}{a_m(i)} \frac{\mathcal{Z}_l(i)}{\mathcal{D}(i)}, \quad l = 1, \ldots, m, \quad (17)$$

donde $\mathscr{D}(i)=\Delta_{i+1}(v_1, v_2, \ldots, v_m)$, y

$$\mathscr{D}_l(i)=$$

$$=\begin{vmatrix} v_1(i+1) & \ldots & v_{l-1}(i+1) & v_{l+1}(i+1) & \ldots & v_m(i+1) \\ v_1(i+2) & \ldots & v_{l-1}(i+2) & v_{l+1}(i+2) & \ldots & v_m(i+2) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ v_1(i+m-1) & \ldots & v_{l-1}(i+m-1) & v_{l+1}(i+m-1) & \ldots & v_m(i+m-1) \end{vmatrix},$$

es decir, $\mathscr{D}_l(i)$ se obtiene del determinante $\mathscr{D}(i)$ excluyendo su l-ésima columna y la última fila.

Las igualdades (17) son las ecuaciones de diferencias de primer orden respecto a las funciones $c_l(i)$, $l=1, 2, \ldots, m$. Ya que $c_l(i)$ puede ser determinada con exactitud hasta una constante, entonces de (17) hallamos la representación explícita para $c_l(i)$:

$$c_l(i)=\sum_{j=i_0}^{i-1}(-1)^{m+l}\frac{f(j)}{a_m(j)}\frac{\mathscr{D}_l(j)}{\mathscr{D}(j)}, \quad l=1, 2, \ldots, m.$$

Sustituyendo esta expresión en (13) y cambiando el orden de sumación en la representación obtenida, tendremos la siguiente fórmula para una solución particular $\bar{y}(i)$ de la ecuación no homogénea (7):

$$\bar{y}(i)=\sum_{l=1}^{m}c_l(i)\,v_l(i)=$$

$$=\sum_{j=i_0}^{i-1}\left[f(j)\sum_{l=1}^{m}(-1)^{m+l}\mathscr{D}_l(j)\,v_l(i)\right]\Big/(\mathscr{D}(j)\,a_m(j))=$$

$$=\sum_{j=i_0}^{i-1}G(i, j)\,f(j),$$

donde

$$G(i, j)=\frac{1}{\mathscr{D}(j)\,a_m(j)}\sum_{k=1}^{m}(-1)^{m+k}\mathscr{D}_k(j)\,v_k(i). \qquad (18)$$

Notemos, que la suma que hay en (18) se calcula fácilmente

$$\sum_{k=1}^{m} (-1)^{m+k} \mathscr{D}_k(j)\, v_k(i) =$$

$$= \begin{vmatrix} v_1(j+1) & v_2(j+1) & \dots & v_m(j+1) \\ v_1(j+2) & v_2(j+2) & \dots & v_m(j+2) \\ \dots & \dots & \dots & \dots \\ v_1(j+m-1) & v_2(j+m-1) & \dots & v_m(j+m-1) \\ v_1(i) & v_2(i) & \dots & v_m(i) \end{vmatrix}.$$

Esta suma es igual a cero para $j = i - 1,\ i - 2,\ \dots$ $\dots,\ i - m + 1$. De esta forma, la solución particular de la ecuación (7) tiene la siguiente representación:

$$\bar{y}(i) = \sum_{j=i_0}^{i-m} \frac{\begin{vmatrix} v_1(j+1) & \dots & v_m(j+1) \\ \dots & \dots & \dots \\ v_1(j+m-1) & \dots & v_m(j+m-1) \\ v_1(i) & \dots & v_m(i) \end{vmatrix}}{\begin{vmatrix} v_1(j+1) & \dots & v_1(j+m) \\ \dots & \dots & \dots \\ v_1(j+1) & \dots & v_m(j+m) \end{vmatrix}} \frac{f(j)}{a_m(j)}, \qquad (19)$$

donde i_0 es arbitrario, y para $i = i_0,\ i_0 + 1,\ \dots,\ i_0 + m - 1$ tenemos $\bar{y}(i) = 0$.

Para una ecuación de primer orden ($m = 1$) la fórmula (19) toma la forma siguiente:

$$\bar{y}(i) = \sum_{j=i_0}^{i-1} \frac{v_1(i)}{v_1(j+1)} \cdot \frac{f(j)}{a_1(j)}, \qquad \bar{y}(i_0) = 0. \qquad (20)$$

4. Ejemplos. Examinemos algunos ejemplos que ilustran la aplicación de la teoría general. Supongamos que se exige hallar la solución general de la ecuación de primer orden

$$y(i+1) - e^{2i} y(i) = 6i^2 e^{i^2+i}. \qquad (21)$$

Hallemos primeramente la solución de la ecuación homogénea

$$y(i+1) - e^{2i} y(i) = 0. \qquad (22)$$

De (22) obtenemos sucesivamente

$$y(i+1) = e^{2i} y(i) = e^{2i} e^{2(i-1)} y(i-1) = \dots =$$

$$= e^{2\sum_{h=1}^{i} h}\, y(1) = e^{i(i+1)} y(1).$$

Poniendo aquí $y(1) = 1$ hallaremos una solución particular $v_1(i)$ de la ecuación homogénea (22) en la forma $v_1(i) = e^{i(i-1)}$. Por consiguiente, la solución general de la ecuación homogénea tiene la forma $\bar{\bar{y}}(i) = ce^{i(i-1)}$, donde c es una constante arbitraria.

Construyamos ahora una solución particular de la ecuación no homogénea (21), utilizando la fórmula (20). De (20) obtenemos

$$\bar{y}(i) = \sum_{k=i_0}^{i-1} \frac{e^{i(i-1)}}{e^{k(k+1)}} \cdot \frac{6k^2e^{k^2+k}}{1} = 6e^{i(i-1)} \sum_{k=i_0}^{i-1} k^2.$$

Ya que i_0 puede ser elegido arbitrariamente, entonces, poniendo aquí $i_0 = 1$, tendremos $\bar{y}(i) = i(i-1)(2i-1) \times \times e^{i(i-1)}$. Luego, en virtud del teorema 3 la solución general de la ecuación (21) se escribe en la forma

$$y(i) = \bar{y}(i) + \bar{\bar{y}}(i) = [c + i(i-1)(2i-1)]\, e^{i(i-1)},$$

donde c es una constante arbitraria. El problema está resuelto.

Hallemos ahora la solución general de la ecuación de segundo orden

$$a_2(i)\, y(i+2) + a_1(i)\, y(i+1) + a_0(i)\, y(i) = f(i), \qquad (23)$$

donde $i = 0, 1, 2, \ldots$,

$$a_2(i) = i^2 - i + 1, \quad a_0(i) = a_2(i+1) = i^2 + i + 1,$$
$$a_1(i) = -a_0(i) - a_2(i) = -2(i^2 + 1), \qquad (24)$$
$$f(i) = 2^i(i^2 - 3i + 1) = 2^i[2a_2(i) - a_0(i)].$$

Ya que los coeficientes $a_2(i)$ y $a_0(i)$ son distintos de cero, entonces para hallar la solución general de la ecuación (23) se puede aplicar la teoría general.

Primeramente construyamos soluciones linealmente independientes de la ecuación homogénea. Utilizando (24), es posible escribirla en la siguiente forma:

$$a_2(i)\, y(i+2) - [a_2(i) + a_2(i+1)]\, y(i+1) + a_2(i+1)\, y(i) = 0,$$

o

$$a_2(i)[y(i+2) - y(i+1)] - a_2(i+1)[y(i+1) - y(i)] = 0. \qquad (25)$$

Las soluciones particulares $v_1(i)$ y $v_2(i)$ de la ecuación homogénea (25) destaquemos mediante las siguientes condi-

ciones: $v_1(0) = v_1(1) = 1$, $v_2(0) = 0$, $v_2(1) = 3$. Como el determinante

$$\Delta_0(v_1, v_2) = \begin{vmatrix} v_1(0) & v_1(1) \\ v_2(0) & v_2(1) \end{vmatrix} = 3 \neq 0,$$

entonces en virtud del lema 3 las funciones $v_1(i)$ y $v_2(i)$ serán las soluciones linealmente independientes de la ecuación (25).

Encontremos la forma explícita de $v_1(i)$ y $v_2(i)$. De (25) se deduce inmediatamente que $v_1(i) \equiv 1$. Construyamos $v_2(i)$. De (25) obtenemos sucesivamente

$$y(i+2) - y(i+1) = \frac{a_2(i+1)}{a_2(i)}[y(i+1) - y(i)] =$$

$$= \frac{a_2(i+1)}{a_2(i-1)}[y(i) - y(i-1)] = \ldots \frac{a_2(i+1)}{a_2(0)}[y(1) - y(0)].$$

Teniendo en cuenta los valores iniciales para $y(i)$, de aquí hallamos

$$v_2(i+1) - v_2(i) = 3a_2(i) = 3(i^2 - i + 1). \qquad (26)$$

Sumando los miembros izquierdo y derecho de (26) por i desde cero hasta $k-1$, tendremos

$$v_2(k) = v_2(0) + 3\sum_{i=0}^{k-1}(i^2 - i + 1) = k(k^2 - 3k + 5).$$

Así pues, hemos hallado soluciones particulares de la ecuación homogénea (25)

$$v_1(k) \equiv 1, \quad v_2(k) = k(k^2 - 3k + 5), \qquad (27)$$

y la solución general de (25) tiene la forma

$$\bar{\bar{y}}(k) = c_1 + c_2 k(k^2 - 3k + 5).$$

Construyamos ahora una solución particular de la ecuación no homogénea (23). Sustituyendo (24) y (27) en la

fórmula (19), obtendremos

$$\bar{y}(i) = \sum_{k=0}^{i-2} \frac{v_2(i)-v_2(k+1)}{v_2(k+2)-v_2(k+1)} \cdot \frac{f(k)}{a_2(k)} =$$

$$= \sum_{k=0}^{i-2} \frac{v_2(i)-v_2(k+1)}{3a_2(k+1)\,a_2(k)} \left[2^{k+1}a_2(k)-2^k a_2(k+1)\right] =$$

$$= \frac{1}{3} \sum_{k=0}^{i-2} \left[v_2(i)-v_2(y+1)\right] \left[\frac{2^{k+1}}{a_2(k+1)} - \frac{2^k}{a_2(k)}\right]. \qquad (28)$$

Aquí fue utilizada la igualdad (26).

Calculemos la expresión obtenida. Designando

$$v(k) = v_2(i) - v_2(k+1), \qquad u(k) = \frac{2^k}{a_2(k)},$$

escribamos (28) de la siguiente forma:

$$\bar{y}(i) = \frac{1}{3} \sum_{k=0}^{i-2} \left[u(k+1)-u(k)\right] v(k).$$

Utilicemos ahora la fórmula de sumación por miembros (véase (8) § 1) para el caso de una red uniforme con paso $h = 1$. Esto da

$$\bar{y}(i) = -\frac{1}{3} \sum_{k=0}^{i-1} u(k)\left[v(k)-v(k-1)\right] +$$
$$+ \frac{1}{3}\left[u(i-1)\,v(i-1) - u(0)\,v(-1)\right].$$

Ya que en virtud de (26), de la condición $v_2(0) = 0$ y de la definición de las funciones $v(k)$ y $u(k)$ tenemos

$$v(k) - v(k-1) = v_2(k) - v_2(k+1) = -3a_2(k),$$
$$v(i-1) = v_2(i) - v_2(i) = 0,$$
$$v(-1) = v_2(i) - v_2(0) = v_2(i),$$

entonces

$$\bar{y}(i) = \sum_{k=0}^{i-1} 2^k - \frac{1}{3} v_2(i) = 2^i - 1 - \frac{1}{3} i(i^2 - 3i + 5).$$

Por lo tanto, está hallada la solución particular de (23). En virtud del teorema 3 la solución general de la ecuación

no homogénea de segundo orden (23) tiene la forma

$$y(i) = \bar{y}(i) + \bar{\bar{y}}(i) = 2^i - 1 - \frac{1}{3} i (i^2 - 3i + 5) +$$
$$+ c_1 + c_2 i (i^2 - 3i + 5) + \bar{c}_1 + 2^i + \bar{c}_2 i (i^2 - 3i + 5),$$

dondo $\bar{c}_1 = c_1 - 1$ y $\bar{c}_2 = c_2 - \frac{1}{3}$ son las constantes arbitrarias. El problema está resuelto.

§ 3. Solución de ecuaciones lineales con coeficientes constantes

1. Ecuación característica. Caso de raíces simples. Examinemos ahora la clase importante de ecuaciones de diferencias: las ecuaciones lineales con coeficientes constantes. Para las ecuaciones de esta clase el problema de encontrar soluciones linealmente independientes de las respectivas ecuaciones homogéneas puede ser resuelto bastante sencillamente. Y, como fué mostrado más arriba, a esto se reduce el problema de la solución de una ecuación de diferencias no homogénea.

Ocupémonos en buscar soluciones linealmente independientes de la ecuación lineal homogénea de m-ésimo orden con coeficientes constantes

$$a_m y(i+m) + a_{m-1} y(i+m-1) + \ldots$$
$$\ldots + a_0 y(i) = 0. \qquad (1)$$

Buscaremos soluciones particulares de (1) en la forma $v(i) = q^i$ donde el número q está sujeto a definición posteriormente. Sustituyendo $v(i)$ en lugar de $y(i)$ en (1), obtenemos la ecuación

$$q^i (a_m q^m + a_{m-1} q^{m-1} + \ldots + a_1 q + a_0) = 0.$$

Ya que se busca una solución no idénticamente igual a cero de la ecuación (1), entonces, reduciendo en q^i, obtenemos de aquí la ecuación para q:

$$a_m q^m + a_{m-1} q^{m-1} + \ldots + a_1 q + a_0 = 0. \qquad (2)$$

La ecuación (2) se llama *ecuación característica* para (1). Las raíces $q_1, q_2, \ldots, q_m$ de la ecuación (2) pueden ser tanto simples como múltiples. Examinemos por separado cada caso posible.

Dado las raíces simples. Mostremos que las funciones

$$v_1(i) = q_1^i, \quad v_2(i) = q_2^i, \quad \ldots, \quad v_m(i) = q_m^i \tag{3}$$

son las soluciones linealmente independientes de la ecuación (1).

En efecto, en virtud del lema 3 es suficiente mostrar que al menos para un i el determinante $\Delta_i(v_1, v_2, \ldots, v_m) \neq 0$. Poniendo $i = 0$, hallaremos

$$\Delta_0(v_1, \ldots, v_m) =$$

$$= \begin{vmatrix} 1 & q_1 & q_1^2 & \ldots & q_1^{m-1} \\ 1 & q_2 & q_2^2 & \ldots & q_2^{m-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & q_m & q_m^2 & \ldots & q_m^{m-1} \end{vmatrix} = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ q_1 & q_2 & \ldots & q_m \\ q_1^2 & q_2^2 & \ldots & q_m^2 \\ \cdot & \cdot & \cdot & \cdot \\ q_1^{m-1} & q_2^{m-1} & \ldots & q_m^{m-1} \end{vmatrix}$$

y, por consiguiente, $\Delta_0(v_1, \ldots, v_m)$ es el determinante de Vandermonde. El es distinto de cero ya que todos los q_k son diferentes. De esta manera, las funciones (3) son realmente las soluciones linealmente independientes de (1) y por eso la solución general de la ecuación homogénea (1) puede ser escrita en la forma

$$y(i) = c_1 q_1^i + c_2 q_2^i + \ldots + c_m q_m^i, \tag{4}$$

donde $c_1, c_2, \ldots, c_m$ son las constantes arbitrarias.

Si las raíces $q_1, q_2, \ldots, q_m$ son reales, entonces la solución real $y(i)$ destaca mediante la elección de las constantes $c_1, c_2, \ldots, c_m$ que sean números reales. Examinemos ahora el problema de separar la solución real, si entre las raíces hay raíces complejas.

Sea $q_n = \rho(\cos\varphi + i^* \operatorname{sen}\varphi)$, $(i^* = \sqrt{-1})$ una raíz compleja de la ecuación característica (2). Entonces existe la raíz $q_s = \rho(\cos\varphi - i^* \operatorname{sen}\varphi)$ de la ecuación (2), conjugada a q_n. Examinemos la parte de la solución general (4) formada por la combinación lineal de q_n^i y q_s^i:

$$y(i) = c_n q_n^i + c_s q_s^i =$$
$$= \rho\,[(c_n + c_s)\cos i\varphi + i^*(c_n - c_s)\operatorname{sen} i\varphi].$$

La función $y(i)$ tendrá valores reales, si las constantes c_n y c_s son complejo conjugadas. Poniendo $c_n = 0{,}5(\bar{c}_n - i^*\bar{c}_s)$, $c_s = 0{,}5(\bar{c}_n + i^*\bar{c}_s)$, donde $\bar{c}_n$ y $\bar{c}_s$ son los números

enteros arbitrarios, obtendremos $y(i) = \rho^i = (\overline{c}_n \cos i\varphi + \overline{c}_s \operatorname{sen} i\varphi)$.

2. **Caso de raíces múltiples.** Supongamos ahora que la ecuación característica (2) tiene la raíz q_1 de multiplicidad n_1, q_2 de multiplicidad n_2, etc., es decir $q_1, q_2, \ldots, q_s$ son raíces diferentes de multiplicidad $n_1, n_2, \ldots, n_s$ respectivamente, $n_1 + n_2 + \ldots + n_s = m$. Construyamos soluciones linealmente independientes de la ecuación (1). Nosotros necesitaremos el siguiente lema:

LEMA 4. *Si q_l es la raíz de la ecuación característica* (2) *de multiplicidad n_l, entonces son válidas las igualdades*

$$\sum_{k=0}^{m} a_k k^p q_l^k = 0, \quad p = 0, \quad 1, \ldots, n_l - 1. \tag{5}$$

En efecto, como q_l es la raíz de multiplicidad n_l de la ecuación (2), entonces tienen lugar las igualdades

$$\sum_{k=0}^{m} a_k q_l^k = 0, \tag{6}$$

$$\sum_{k=0}^{m} k(k-1) \ldots (k-s+1) a_k q_l^k = 0,$$

$$s = 1, 2, \ldots, n_l - 1, \tag{7}$$

obtenidas de (2) diferenciando s veces y multiplicando el resultado por q_l^s. Mostremos que la igualdad (5) es equivalente a (6) y (7). Es evidente, que es necesario demostrar solamente la equivalencia de (7) y (5) para $p \geqslant 1$.

Ya que $P_s(k) = k(k-1) \ldots (k-s+1)$ es un polinomio de k de grado s, entonces multiplicando (5) por el coeficiente correspondiente del polinomio $P_s(k)$ para $p = 1, 2, \ldots, s$ y sumando las igualdades obtenidas, tendremos las relaciones (7).

Mostremos ahora, que de (7) se deducen las igualdades (5) para $p = 1, 2, \ldots, n_l - 1$. Utilicemos el desarrollo para k^p:

$$k^p = \sum_{s=1}^{p} k(k-1) \ldots (k-s+1) \alpha_s, \quad 1 \leqslant p \leqslant k, \tag{8}$$

donde $\alpha_s = \alpha_s(p)$ será indicado más abajo. Multipliquemos la s-ésima igualdad (7) por α_s y sumemos por s desde 1

hasta p. En virtud de (8) obtenemos

$$0=\sum_{s=1}^{p}\alpha_s\left(\sum_{k=0}^{m}k(k-1)\dots(k-s+1)a_kq_l^k\right)=$$

$$=\sum_{k=0}^{m}a_kq_l^k\left(\sum_{s=1}^{p}k(k-1)\dots(k-s+1)\alpha_s\right)=\sum_{k=0}^{m}a_kk^pq_l^k.$$

Queda por fundamentar el desarrollo (8). Observemos, que a la izquierda y a la derecha en (8) hay polinomios de k de p-ésimo grado. Si ponemos $\alpha_p=1$, entonces los coeficientes de la mayor potencia de k a la izquierda y a la derecha en (8) serán iguales y los coeficientes de la menor potencia son iguales a cero.

Hallemos $\alpha_1, \alpha_2, \dots, \alpha_{p-1}$ igualando los valores de los polinomios en $p-1$ puntos distintos, por ejemplo, poniendo $k=1, 2, \dots, p-1$. Para $k=1$ esto da $\alpha_1=1$. Cuando $k=n$, $2\leqslant n\leqslant p-1$, tendremos $n^p=$

$$=\sum_{s=1}^{p}n(n-1)\dots(n-s+1)\alpha_s=\sum_{s=1}^{n}n(n-1)\dots$$

$$\dots(n-s+1)\alpha_s=n!\,\alpha_n+n!\sum_{s=1}^{n-1}\frac{\alpha_s}{(n-s)!}.$$

De aquí se puede hallar α_n, si ya están determinados α_1, α_2, $\dots$, α_{n-1}. De esta forma obtenemos la siguiente fórmula recurrente para encontrar los coeficientes α_n:

$$\alpha_n=\frac{n^p}{n!}-\sum_{s=1}^{n-1}\frac{\alpha_s}{(n-s)!},\quad n=2,\ 3,\ \dots,\ p-1,\ \alpha_1=1.$$

El lema está demostrado.

Utilizando el lema 4, hallaremos m soluciones particulares de la ecuación homogénea (1). Ya que es válida la igualdad

$$(j+k)^n=\sum_{p=0}^{n}C_n^pk^pj^{n-p},\quad C_n^p=\frac{n!}{p!\,(n-p)!},$$

entonces, multiplicando (5) por $C_n^pj^{n-p}q_l^j$ y sumando por p desde cero hasta $n\leqslant n_l-1$, obtendremos que para cualquier j tienen lugar las igualdades

$$\sum_{k=0}^{m}a_k(j+k)^nq_l^{k+j}=0,\quad n=0,\ 1,\ \dots,\ n_l-1.$$

Utilizando estas igualdades, se halla fácilmente que las funciones reticulares

$$v_{n_1+n_2+\dots+n_{l-1}+n+1}(j) = j^n q_l^j,$$
$$0 \leqslant n \leqslant n_l - 1, \quad l = 1, 2, \dots, s, \tag{9}$$

son las soluciones particulares de la ecuación homogénea (1), es decir, si q_l es la raíz de multiplicidad n_l de la ecuación característica, entonces las funciones

$$q_l^j, \; jq_l^j, \; \dots, \; j^{n_l-1}q_l^j, \quad l = 1, 2, \dots, s$$

son las soluciones de la ecuación (1).

Queda por mostrar, que las funciones $v_1(j), \dots, v_m(j)$, definidas en (9), son las soluciones linealmente independientes. Para eso calculemos el determinante $\Delta_0(v_1, \dots, v_m)$, el cual en el caso dado posee la forma

$$\Delta_0(v_1, \dots, v_m) =$$
$$= \begin{vmatrix} 1 & q_1 & q_1^2 & \dots & q_1^k & \dots & q_1^{m-1} \\ 0 & q_1 & 2q_1^2 & \dots & kq_1^k & \dots & (m-1)q_1^{m-1} \\ 0 & q_1 & 2^2q_1^2 & \dots & k^2q_1^k & \dots & (m-1)^2q_1^{m-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & q_2 & q_2^2 & \dots & q_2^k & \dots & q_2^{m-1} \\ 0 & q_2 & 2q_2^2 & \dots & kq_2^k & \dots & (m-1)q_2^{m-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & q_s & 2^{n_s-1}q_s^2 & \dots & k^{n_s-1}q_s^k & \dots & (m-1)^{n_s-1}q_s^{m-1} \end{vmatrix}$$

El puede ser obtenido directamente del determinante de Vandermonde

$$W(x_1, x_2, \dots, x_m) =$$
$$= \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{m-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{m-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{m-1} & x_{m-1}^2 & \dots & x_{m-1}^{m-1} \\ 1 & x_m & x_m^2 & \dots & x_m^{m-1} \end{vmatrix} = \prod_{i=1}^{m-1} \prod_{j=i+1}^{m} (x_j - x_i)$$

de la siguiente forma. Tomemos la primera derivada de W por x_2 y multipliquémosla por x_2. Denotemos el resul-

tado mediante $W'_2 = x_2 \frac{\partial W}{\partial x_2}$. A continuación calculemos

$$W'_3 = x_3 \frac{\partial}{\partial x_3}\left(x_3 \frac{\partial W_2}{\partial x_3}\right), \quad W'_4 = x_4 \frac{\partial}{\partial x_4} \times$$

$$\times \left(x_4 \frac{\partial}{\partial x_4}\left(x_4 \frac{\partial W_3}{\partial x_4}\right)\right), \ \ldots$$

etc., hasta que obtengamos W_{n_1}. Después calculemos $W'_{n_1+2} = x_{n_1+2} \frac{\partial W'_{n_1}}{\partial x_{n_1+2}}$ y continuemos el proceso de diferenciación, calculando $W'_{n_1+3} = x_{n_1+3} \frac{\partial}{\partial x_{n_1+3}}\left(x_{n_1+3} \frac{\partial W'_{n_1+2}}{\partial x_{n_1+3}}\right)$, hasta que obtengamos $W_{n_1+n_2}$, etc. Como resultado obtendremos $W_m = W_m(x_1, x_2, \ldots, x_m)$. Pongamos aquí $x_1 = x_2 = \ldots = x_{n_1} = q_1$, $x_{n_1+1} = x_{n_1+2} = \ldots = x_{n_1+n_2} = q_2$, etc. Es fácil convencerse de que $\Delta_0(v_1, v_2, \ldots, v_m) = W_m$ y cálculos sencillos dan

$$W'_m = \prod_{h=1}^{s} \prod_{m=1}^{n_h-1} m! q_h^m \prod_{i=1}^{s-1} \prod_{j=i+1}^{s} (q_j - q_i)^{n_i n_j}.$$

De aquí se deduce que $\Delta_0(v_1, \ldots, v_m) \neq 0$, ya que $q_j \neq q_i$ cuando $j \neq i$ y por eso las funciones $v_1(j)$, $v_2(j)$, ..., $v_m(j)$, construidas más arriba, son las soluciones linealmente independientes de la ecuación homogénea (1). Además, la solución general de la ecuación (1) se escribe en la forma

$$y(j) = \sum_{l=1}^{s} \sum_{n=0}^{n_l-1} c_n^{(l)} j^n q_l^j,$$

donde $c_n^{(l)}$ son las constantes arbitrarias.

3. **Ejemplos.** Examinemos los ejemplos más simples de encontrar la solución general de una ecuación en diferencias homogénea con coeficientes constantes.

1. Se exige hallar la solución general de la ecuación

$$y(i+2) - y(i+1) - 2y(i) = 0. \qquad (10)$$

Formamos la ecuación característica $q^2 - q - 2 = 0$ y encontramos sus raíces $q_1 = 2$ y $q_2 = -1$. Puesto que las raíces son simples, la solución general de la ecuación (10) tiene la forma

$$y(i) = c_1 2^i + c_2 (-1)^i.$$

2. Hallar la solución general de la ecuación de cuarto orden

$$y(j+4)-2y(j+3)+3y(j+2)+2y(j+1)-$$
$$-4y(j)=0. \qquad (11)$$

La ecuación característica $q^4-2q^3+3q^2+2q-4=0$ posee dos raíces reales $q_1=1$, $q_2=-1$ y dos raíces complejas conjugadas $q_3=2\left(\cos\frac{\pi}{3}+i\operatorname{sen}\frac{\pi}{3}\right)$ y $q_4=2\times\left(\cos\frac{\pi}{3}-i\operatorname{sen}\frac{\pi}{3}\right)$, donde $i=\sqrt{-1}$. Por consiguiento, la solución general de la ecuación (11) que toma valores reales, tiene la forma

$$y(j)=c_1+c_2(-1)^j+2^j\left(c_3\cos\frac{\pi}{3}j+c_4\operatorname{sen}\frac{\pi}{3}j\right).$$

3. Hallar la solución general de la ecuación de cuarto orden

$$y(j+4)-7y(j+3)+18y(j+2)-$$
$$-20y(j+1)+8y(j)=0. \qquad (12)$$

La ecuación característica

$$q^4-7q^3+18q^2-20q+8=(q-2)^3(q-1)=0$$

tiene la raíz $q_1=2$ de multiplicidad 3 y la raíz $q_2=1$ de multiplicidad 1. Por lo tanto, la solución general de (12) tiene la forma

$$y(j)=c_1+2^j(c_2+c_3j+c_4j^2),$$

y las funciones reticulares $v_1(j)=1$, $v_2(j)=2^j$, $v_3(j)=j2^j$ y $v_4(j)=j^22^j$ son las soluciones particulares linealmente independientes de (12).

4. Hallar la solución general de la ecuación de cuarto orden

$$y(j+4)+8y(j+2)+16y(j)=0. \qquad (13)$$

La ecuación característica $q^4+8q^2+16=(q^2+4)^2=0$ tiene la raíz compleja $q_1=2\left(\cos\frac{\pi}{2}+i\operatorname{sen}\frac{\pi}{2}\right)$ de multiplicidad 2 y la raíz conjugada a ésta $q_2=2\left(\cos\frac{\pi}{2}-\right.$ $\left.-i\operatorname{sen}\frac{\pi}{2}\right)$, también de multiplicidad 2. Por eso la solu-

ción general de (13) que toma valores reales, tiene la forma

$$y(j) = (c_1 + c_2 j)\, 2^j \cos \frac{\pi}{2} j + (c_3 + c_4 j)\, 2^j \operatorname{sen} \frac{\pi}{2} j.$$

Examinemos dos ejemplos más. En un ejemplo hallaremos la solución del problema de Cauchy para la ecuación no homogénea de primer orden y en el otro hallaremos la solución de un problema de contorno para la ecuación homogénea de cuarto orden.

5. Hallar la solución del siguiente problema:

$$y(i+1) - ay(i) = f(i), \quad i \geqslant 0, \quad y(0) = y_0, \quad (14)$$

donde $a = \text{const}$. La ecuación característica $q - a = 0$ tiene la única raíz $q_1 = a$. Por eso la solución general de la ecuación homogénea tiene la forma $\overset{=}{y}(i) = ca^i$, $c = \text{const}$. Hallaremos una solución particular de la ecuación no homogénea (14), utilizando el método de variación de la constante. La fórmula (20) del § 2 da la siguiente solución particular de la ecuación (14):

$$\bar{y}(i) = \sum_{k=0}^{i-1} a^{i-k-1} f(k) = \sum_{k=0}^{i-1} a^k f(i-k-1).$$

En virtud del teorema 3 la solución general de la ecuación no homogénea (14) tiene la forma

$$y(i) = ca^i + \sum_{k=0}^{i-1} a^k f(i-k-1).$$

Suponiendo aquí $i = 0$, obtenemos (en este caso la suma desaparece) $y_0 = y(0) = c$. De esta forma, la solución del problema (14) viene dada por la fórmula

$$y(i) = y_0 a^i + \sum_{k=0}^{i-1} a^k f(i-k-1), \quad i \geqslant 0.$$

6. Hallemos ahora la solución de la ecuación de cuarto orden

$$y(j+2) - y(j+1) + 2y(j) - y(j-1) + y(j + {} \\ {} + 2) = 0, \quad 2 \leqslant j \leqslant N-2, \quad (15)$$

que satisface las siguientes condiciones de contorno:

$$\begin{aligned} &2y(2) - y(1) + y(0) = 2, \\ &y(3) - y(2) + y(1) - y(0) = 0, \\ &y(N-3) - y(N-2) + y(N-1) - y(N) = 0, \quad (16) \\ &2y(N-2) - y(N-1) + y(N) = 0. \end{aligned}$$

La ecuación característica

$$q^4 - q^3 + 2q^2 - q + 1 = (q^2 - q + 1)(q^2 + 1) = 0,$$

correspondiente a (15), posee las raíces complejas $q_1 = \cos\frac{\pi}{3} + i\,\text{sen}\,\frac{\pi}{3}$, $q_2 = \cos\frac{\pi}{3} - i\,\text{sen}\,\frac{\pi}{3}$, $q_3 = \cos\frac{\pi}{2} + i\,\text{sen}\,\frac{\pi}{2}$ y $q_4 = \cos\frac{\pi}{2} - i\,\text{sen}\,\frac{\pi}{2}$, siendo $i = \sqrt{-1}$. Por consiguiente, la solución general de la ecuación homogénea (15) que toma valores reales, tiene la forma

$$y(j) = c_1 \cos\frac{1}{3}\pi j + c_2\,\text{sen}\,\frac{1}{3}\pi j + c_3 \cos\frac{1}{2}\pi j + c_4\,\text{sen}\,\frac{1}{2}\pi j. \qquad (17)$$

Separemos ahora de la solución general (17), aquella solución que satisface las condiciones de contorno (16). Para eso sustituyamos (17) en (16) y obtendremos el siguiente sistema para las constantes c_1, c_2, c_3 y c_4:

$$\begin{aligned}
&\cos\frac{2\pi}{3}c_1 + \text{sen}\,\frac{2\pi}{3}c_2 - c_3 - c_4 = 2,\\
&c_1 + 0\cdot c_2 + 0\cdot c_3 + 0\cdot c_4 = 0,\\
&\cos\frac{N\pi}{3}c_1 + \text{sen}\,\frac{N\pi}{3}c_2 + 0\cdot c_3 + 0\cdot c_4 = 0,\\
&\cos\frac{(N-2)\pi}{3}c_1 + \text{sen}\,\frac{(N-2)\pi}{3}c_2 - \left(\cos\frac{\pi N}{2} + \text{sen}\,\frac{\pi N}{2}\right)c_3 + \\
&\qquad + \left(\cos\frac{\pi N}{2} - \text{sen}\,\frac{\pi N}{2}\right)c_4 = 0.
\end{aligned}$$

El determinante de este sistema es igual a $-2\,\text{sen}\,\frac{N\pi}{3}\cos\frac{N\pi}{2}$ y es distinto de cero si N es par, pero no múltiplo de 3.

En este caso, teniendo en cuenta la paridad de N, obtenemos $c_1 = c_2 = 0$, $c_3 = c_4 = -1$. De esta forma, si N es par y no es múltiplo de 3, entonces existe la solución del problema de contorno (15), (16) y se da por la fórmula

$$y(j) = -\cos\frac{\pi j}{2} - \text{sen}\,\frac{\pi j}{2}, \quad 0 \leqslant j \leqslant N.$$

Si N es impar o múltiplo de 3, entonces la solución del problema (15), (16) o bien no existe o no es única. Este ejemplo ilustra la diferencia entre los problemas de contorno cuya solución no existe siempre, y el problema de Cauchy que posee solución única.

§ 4. Ecuaciones de segundo orden con coeficientes constantes

1. Solución general de una ecuación homogénea. El presente párrafo está dedicado a las ecuaciones en diferencias de segundo orden con coeficientes constantes

$$a_2 y(j+2) + a_1 y(j+1) + a_0 y(j) = \\ = f(j), \quad a_0, \quad a_2 \neq 0. \qquad (1)$$

Primeramente hallaremos la solución general de la correspondiente ecuación homogénea

$$a_2 y(j+2) + a_1 y(j+1) + a_0 y(j) = 0. \qquad (2)$$

La ecuación característica $a_2 q^2 + a_1 q + a_0 = 0$ posee as raíces

$$q_1 = \frac{-a_1 + \sqrt{a_1^2 - 4a_0 a_2}}{2a_2}, \quad q_2 = \frac{-a_1 - \sqrt{a_1^2 - 4a_0 a_2}}{2a_2}.$$

De acuerdo con la teoría general de ecuaciones de diferencias con coeficientes constantes expuesta en el § 3, las funciones $v_1(j) = q_1^j$ y $v_2(j) = q_2^j$ si $a_1^2 \neq 4a_0 a_2$ y $v_1(j) = q_1^j$, $v_2(j) = j q_1^j$ si $a_1^2 = 4a_0 a_2$, son las soluciones linealmente independientes de la ecuación (2). A continuación nos será cómodo utilizar otras soluciones linealmente independientes

$$v_1(j) = \frac{q_2 q_1^j - q_1 q_2^j}{q_2 - q_1}, \quad v_2(j) = \frac{q_2^j - q_1^j}{q_2 - q_1}, \qquad (3)$$

que toman para $j = 0$ y $j = 1$, los siguientes valores:

$$v_1(0) = 1, \quad v_1(1) = 0, \quad v_2(0) = 0, \quad v_2(1) = 1. \qquad (4)$$

Obviamente, es necesario mostrar tan sólo que las funciones (3) para el caso $a_1^2 = 4a_0 a_2$ son las soluciones de la ecuación homogénea. La independencia lineal de las funciones (3) construidas, se deduce de la condición $\Delta_0(v_1, v_2) \neq \neq 0$, donde

$$\Delta_0(v_1, v_2) = \begin{vmatrix} v_1(0) & v_1(1) \\ v_2(0) & v_2(1) \end{vmatrix}.$$

Pasando al límite en (3) para q_2 que se tiende a q_1, obtendremos las funciones $v_1(j) = -(j-1) q_1^j$ y $v_2(j) = j q_1^{j-1}$, que son realmente las soluciones de la ecuación homogénea (2). Notemos que las funciones $v_1(j)$ y $v_2(j)$ de (3) toman valores reales también en el caso cuando las raíces q_1 y q_2 son com-

plejas. Esto permite no examinar por separado el caso de raíces complejas. Así, la solución general de la ecuación homogénea (2) puede ser escrita en la forma

$$\bar{\bar{y}}(j) = c_1 v_1(j) + c_2 v_2(j) = c_1 \frac{q_2 q_1^j - q_1 q_2^j}{q_2 - q_1} + c_2 \frac{q_2^j - q_1^j}{q_2 - q_1}, \qquad (5)$$

donde c_1 y c_2 son las constantes arbitrarias. Observemos que en virtud de (4), tendremos $\bar{\bar{y}}(0) = c_1$ y $\bar{\bar{y}}(1) = c_2$.

Examinemos un ejemplo. Se exige hallar la solución general de la ecuación homogénea

$$y(j+2) - 2xy(j+1) + y(j) = 0, \qquad (6)$$

donde x es un parámetro que toma cualesquiera valores reales. En este caso tenemos

$$q_1 = x + \sqrt{x^2 - 1}, \quad q_2 = \frac{1}{q_1}, \quad q_2 - q_1 = -2\sqrt{x^2 - 1}. \qquad (7)$$

Sustituyendo (7) en (5), obtendremos la solución general de la ecuación (6) para cualquier x en la forma

$$y(j) = -\frac{(x+\sqrt{x^2-1})^{j-1} - (x+\sqrt{x^2-1})^{-(j-1)}}{2\sqrt{x^2-1}}\, y(0) + \\ + \frac{(x+\sqrt{x^2-1})^{j} - (x+\sqrt{x^2-1})^{-j}}{2\sqrt{x^2-1}}\, y(1). \qquad (8)$$

En particular, si $|x| \leqslant 1$, entonces la fórmula (8) puede ser escrita en la forma

$$y(j) = -\frac{\operatorname{sen}(j-1)\arccos x}{\operatorname{sen}\arccos x}\, y(0) + \frac{\operatorname{sen} j \arccos x}{\operatorname{sen}\arccos x}\, y(1). \qquad (9)$$

(Para obtener (9) fue utilizada la identidad $x = \cos(\arccos x)$)

Aprovechemos el resultado obtenido para resolver el problema proporcionado en el punto 4 del § 1, sobre el cálculo de las integrales

$$I_k(\varphi) = \int_0^{\pi} \frac{\cos k\psi - \cos k\varphi}{\cos\psi - \cos\varphi}\, d\psi, \quad k = 0, 1, \ldots$$

Allí fue mostrado, que este problema se reduce a resolver el problema de Cauchy para la ecuación.

$$I_{k+1} - 2\cos\varphi I_k + I_{k-1} = 0, \quad I_0 = 0, \quad I_1 = \pi. \qquad (10)$$

Esta ecuación es un caso particular de (6) con $x = \cos\varphi$. Ya que $|x| \leqslant 1$, entonces la solución general de la ecua-

ción (10) se da por la fórmula (9), es decir,

$$I_k = -\frac{\operatorname{sen}(k-1)\varphi}{\operatorname{sen}\varphi} I_0 + \frac{\operatorname{sen} k\varphi}{\operatorname{sen}\varphi} I_1.$$

Sustituyendo aquí los valores iniciales para I_k, obtenemos la solución del problema proporcionado

$$I_k(\varphi) = \frac{\pi \operatorname{sen} k\varphi}{\operatorname{sen}\varphi}.$$

En calidad de un segundo ejemplo, examinemos la solución del problema de contorno

$$\begin{gathered} y(j+1) - y(j) + y(j-1) = 0, \quad 1 \leqslant j \leqslant N-1, \\ y(0) = 1, \quad y(N) = 0 \end{gathered} \tag{11}$$

La ecuación del problema (11) es también un caso particular de (6) correspondiente al valor $x = \frac{1}{2}$. La fórmula (9) da la siguiente solución general de la ecuación (11):

$$y(j) = \left(c_1 \operatorname{sen}\frac{(j-1)\pi}{3} + c_2 \operatorname{sen}\frac{j\pi}{3}\right) \Big/ \operatorname{sen}\frac{\pi}{3}.$$

Las constantes c_1 y c_2 se encuentran de las condiciones de contorno para $y(j)$. Si N no es múltiplo de 3, entonces $c_1 = -1$, $c_2 = \operatorname{sen}\frac{1}{3}\pi(N-1)/\operatorname{sen}\frac{1}{3}\pi N$ y la solución del problema (11) tiene la forma

$$y(j) = \operatorname{sen}\frac{1}{3}(N-j)\pi/\operatorname{sen}\frac{1}{3}N\pi, \quad 0 \leqslant j \leqslant N.$$

Si N es múltiplo de 3, entonces la solución del problema de contorno (11) no existe.

2. **Polinomios de Chebishev.** Regresemos ahora a la ecuación (6). Primeramente examinemos el siguiente problema de Cauchy:

$$\begin{gathered} y(n+2) - 2xy(n+1) + y(n) = 0, \quad n \geqslant 0, \\ y(0) = 1 \quad y(1) = x. \end{gathered} \tag{12}$$

Notemos que de (12) se deduce

$$y(2) = 2xy(1) - y(0) = 2x^2 - 1,$$

$$y(3) = 2xy(2) - y(1) = 4x^3 - 3x,$$

y en general $y(n)$ es un polinomio de x de grado n. Designemos este polinomio por $T_n(x)$. Sustituyendo $T_n(x)$ en lugar de $y(n)$ en (12), obtendremos una relación recurrente

que satisface este polinomio

$$\begin{aligned} &T_{n+2}(x) = 2xT_{n+1}(x) - T_n(x), \quad n \geqslant 0, \\ &T_0(x) = 1, \quad T_1(x) = x, \quad -\infty < x < \infty. \end{aligned} \tag{13}$$

Por otra parte, la solución general de la ecuación (12) se da por la fórmula (8) para todo x. Sustituyendo en (8) los valores iniciales para $y(n)$, tendremos

$$T_n(x) = \frac{(x+\sqrt{x^2-1})^n + (x+\sqrt{x^2-1})^{-n}}{2}. \tag{14}$$

En particular, si $|x| \leqslant 1$, entonces, poniendo aquí $x = \cos(\operatorname{arc}\cos x)$, obtendremos

$$T_n(x) = \cos(n \operatorname{arc}\cos x), \quad |x| \leqslant 1.$$

Así hemos hallado la solución del problema (12). La solución es un polinomio $T_n(x)$ el cual para cualquier x se determina por la fórmula (14) o mediante la fórmula

$$T_n(x) = \begin{cases} \cos(n \operatorname{arc}\cos x), & |x| \leqslant 1, \\ \frac{1}{2}[(x+\sqrt{x^2-1})^n + (x+\sqrt{x^2-1})^{-n}], & |x| \geqslant 1. \end{cases} \tag{15}$$

El polinomio $T_n(x)$ se llama *polinomio de Chebishev de primer género y de grado n*.

Examinemos ahora otro problema de Cauchy para la ecuación (6).

$$\begin{aligned} &y(n+2) - 2xy(n+1) + y(n) = 0, \quad n \geqslant 0 \\ &y(0) = 1, \quad y(1) = 2x. \end{aligned} \tag{16}$$

Es evidente, que aquí también $y(n)$ es un polinomio de x de grado n. Denotémoslo mediante $U_n(x)$. Obtengamos la forma explícita para $U_n(x)$. Sustituyendo los valores iniciales para $y(n)$ en (8), tendremos para cualquier x:

$$\begin{aligned} U_n(x) &= \frac{2x(x+\sqrt{x^2-1})^n - (x+\sqrt{x^2-1})^{n-1}}{2\sqrt{x^2-1}} + \\ &+ \frac{(x+\sqrt{x^2-1})^{-(n-1)} - 2x(x+\sqrt{x^2-1})^{-n}}{2\sqrt{x^2-1}} = \\ &= \frac{(x+\sqrt{x^2-1})^{n+1} - (x+\sqrt{x^2-1})^{-(n+1)}}{2\sqrt{x^2-1}}. \end{aligned} \tag{17}$$

En particular, si $|x| \leqslant 1$, entonces

$$U_n(x) = \frac{\operatorname{sen}(n+1)\arccos x}{\operatorname{sen}\arccos x}$$

El polinomio $U_n(x)$ se llama *polinomio de Chebishev de segundo género y de grado n* y se determina por las fórmulas

$$U_n(x) = \begin{cases} \dfrac{\operatorname{sen}(n+1)\arccos x}{\operatorname{sen}\arccos x}, & |x| \leqslant 1, \\ \dfrac{1}{2\sqrt{x^2-1}}\left[(x\sqrt{x^2-1})^{n+1} - \right. & \\ \left. \qquad -(x+\sqrt{x^2-1})^{-(n+1)}\right], & |x| \geqslant 1 \end{cases} \tag{18}$$

De (16), obtenemos para los polinomios $U_n(x)$ la siguiente relación recurrente:

$$U_{n+2}(x) = 2xU_{n+1}(x) - U_n(x), \quad n \geqslant 0,$$
$$U_0(x) = 1, \quad U_1(x) = 2x. \tag{19}$$

La fórmula (17) permite obtener en lugar de (8) la siguiente representación para la solución general de la ecuación (6):

$$y(n) = -c_1U_{n-2}(x) + c_2U_{n-1}(x).$$

Obtengamos una representación más para la solución general de la ecuación (6). Mostremos que las funciones $v_1(n) = T_n(x)$ y $v_2(n) = U_{n-1}(x)$ son las soluciones linealmente independientes de la ecuación homogénea (6). En realidad, necesitamos probar solamente su independencia lineal. Ya que el determinante

$$\Delta_0(v_1, v_2) = \begin{vmatrix} T_0(x) & T_1(x) \\ U_{-1}(x) & U_0(x) \end{vmatrix} = \begin{vmatrix} 1 & x \\ 0 & 1 \end{vmatrix} = 1$$

es distinto de cero, entonces se cumple la afirmación. Por consiguiente, la solución general de la ecuación (6) se puede representar en la forma

$$y(n) = c_1T_n(x) + c_2U_{n-1}(x), \tag{20}$$

donde c_1 y c_2 son las constantes arbitrarias, y las funciones $T_n(x)$ y $U_n(x)$ se determinan, para cualesquiera x y n, por las fórmulas (14) y (17).

Como conclusión citemos algunas relaciones fácilmente comprobables, que expresan la relación entre los polinomios de Chebishev $T_n(x)$ y $U_n(x)$ y además las propiedades

de estos polinomios. Tienen lugar las siguientes fórmulas:

$$T_n(x) = T_{-n}(x), \quad U_{-n}(x) = -U_{n-2}(x), \quad n \geqslant 0, \tag{21}$$

$$T_{in}(x) = T_i(T_n(x)), \quad U_{in-1}(x) = U_{i-1}(T_n(x))\, u_{n-1}(x), \tag{22}$$

$$T_{2n}(x) = 2(T_n(x))^2 - 1, \tag{23}$$

$$T_{n-1}(x) - xT_n(x) = (1 - x^2)\, U_{n-1}(x), \tag{24}$$

$$U_{n-1}(x) - xU_n(x) = -T_{n+1}(x), \tag{25}$$

$$U_{n+i}(x) + U_{n-i}(x) = 2T_i(x)\, U_n(x). \tag{26}$$

De (26) para el correspondiente cambio de índices i y n, obtenemos

$$U_{n+i-1}(x) + U_{n-i-1}(x) = 2T_i(x)\, U_{n-1}(x), \tag{27}$$

$$U_{n+i}(x) + U_{n-i-2}(x) = 2T_{i+1}(x)\, U_{n-1}(x). \tag{28}$$

Poniendo en (26)—(28) $i = n$, tendremos

$$2T_n(x)\, U_n(x) = U_{2n}(x) + 1, \tag{29}$$

$$2T_n(x)\, U_{n-1}(x) = U_{2n-1}(x), \tag{30}$$

$$2T_{n+1}(x)\, U_{n-1}(x) = U_{2n}(x) - 1. \tag{31}$$

Aquí se tuvieron en cuenta las igualdades (21) y que $U_0(x) = 1$, $U_{-1}(x) = 0$. Si ponemos $n = 0$ en (26), entonces obtendremos

$$2T_n(x) = U_n(x) - U_{n-2}(x). \tag{32}$$

3. Solución general de una ecuación no homogénea. Construyamos ahora la solución general de la ecuación no homogénea (1)

$$a_2 y(n+2) + a_1 y(n+1) + a_0 y(n) = f(n). \tag{33}$$

En virtud del teorema 3 la solución general de la ecuación (33) es la suma $y(n) = \overline{y}(n) + \overline{\overline{y}}(n)$, donde $\overline{\overline{y}}(n)$ es la solución general de la ecuación homogénea (2), e $\overline{y}(n)$ es la solución particular de la ecuación no homogénea (33).

Más arriba fue mostrado, que las funciones

$$v_1(n) = \frac{q_2 q_1^n - q_1 q_2^n}{q_2 - q_1}, \quad v_2(n) = \frac{q_2^n - q_1^n}{q_2 - q_1}, \tag{34}$$

son las soluciones linealmente independientes de la ecuación (2), y que la solución $\overline{y}(n)$ se determina por la fórmu-

la (5):

$$\bar{\bar{y}}(n) = c_1 v_1(n) + c_2 v_2(n).$$

Para hallar una solución particular $\bar{y}(n)$ de la ecuación (33) utilicemos el método de variación de las constantes, expuesto en el punto 3 del § 2. La fórmula (19) del § 2 da la solución $\bar{y}(n)$ en la siguiente forma:

$$\bar{y}(n) = \sum_{k=n_0}^{n-1} \frac{\begin{vmatrix} v_1(k+1) & v_2(k+1) \\ v_1(k) & v_2(k) \end{vmatrix}}{\begin{vmatrix} v_1(k+1) & v_1(k+2) \\ v_2(k+1) & v_2(k+2) \end{vmatrix}} \cdot \frac{f(k)}{a_2}.$$

Como resultado de cálculos no complicados tendremos

$$\bar{y}(n) = \sum_{k=n_0}^{n-2} \frac{q_2^{n-k-1} - q_1^{n-k-1}}{q_2 - q_1} \cdot \frac{f(k)}{a_2}, \quad n \neq n_0, \quad n_0 + 1$$

y

$$\bar{y}(n_0) = \bar{y}(n_0 + 1) = 0.$$

Por lo tanto, la solución general de la ecuación no homogénea (33) tiene la forma

$$y(n) = c_1 \frac{q_2 q_1^n - q_1 q_2^n}{q_2 - q_1} + c_2 \frac{q_2^n - q_1^n}{q_2 - q_1} + \sum_{k=n_0}^{n-2} \frac{q_2^{n-k-1} - q_1^{n-k-1}}{q_2 - q_1} \cdot \frac{f(k)}{a_2}, \qquad (35)$$

donde c_1 y c_2 son las constantes arbitrarias.

Si se resuelve el problema de Cauchy, es decir, si se busca la solución de la ecuación (33) que satisface las condiciones

$$y(n_0) = y_0, \quad y(n_0 + 1) = y_1, \qquad (36)$$

entonces de (35) y (36) obtendremos la siguiente representación para resolver este problema:

$$y(n) = y_0 \frac{q_2 q_1^{n-n_0} - q_1 q_2^{n-n_0}}{q_2 - q_1} + y_1 \frac{q_2^{n-n_0} - q_1^{n-n_0}}{q_2 - q_1} + \sum_{k=n_0}^{n-2} \frac{q_2^{n-k-1} - q_1^{n-k-1}}{q_2 - q_1} \cdot \frac{f(k)}{a_2}. \qquad (37)$$

Hallemos ahora la solución del primer problema de contorno para la ecuación de diferencias de segundo orden con coeficientes constantes. Será cómodo escribir dicho problema en la siguiente forma:

$$a_2 y(n+1) + a_1 y(n) + a_0 y(n-1) = -f(n),$$
$$1 \leqslant n \leqslant N-1, \qquad (38)$$
$$y(0) = \mu_1, \quad y(N) = \mu_2.$$

Esta escritura se diferencia de (33) por el desplazamiento del índice n, por tanto, empleando (35), obtenemos la siguiente fórmula para la solución general de la ecuación (38):

$$y(n) = c_1 \frac{q_2 q_1^n - q_1 q_2^n}{q_2 - q_1} + c_2 \frac{q_2^n - q_1^n}{q_2 - q_1} -$$
$$- \sum_{k=1}^{n-1} \frac{q_2^{n-k} - q_1^{n-k}}{q_2 - q_1} \cdot \frac{f(k)}{a_2}. \qquad (39)$$

Definamos las constantes c_1 y c_2 de la condición de que la solución (39) tome para $n = 0$ y $n = N$ los valores prefijados $y(0) = \mu_1$ e $y(N) = \mu_2$. Omitiendo cálculos no complicados, obtendremos la siguiente fórmula para la solución del problema de contorno (38):

$$y(n) = \frac{(q_1 q_2)^n (q_2^{N-n} - q_1^{N-n})}{q_2^N - q_1^N} \mu_1 + \frac{q_2^n - q_1^n}{q_2^N - q_1^N} \mu_2 +$$
$$+ \sum_{k=1}^{n-1} \frac{(q_1 q_2)^{n-k} (q_2^{N-n} - q_1^{N-n}) (q_2^k - q_1^k)}{(q_2 - q_1)(q_2^N - q_1^N)} \cdot \frac{f(k)}{a_2} +$$
$$+ \sum_{k=n}^{N-1} \frac{(q_2^{N-k} - q_1^{N-k}) (q_2^n - q_1^n)}{(q_2 - q_1)(q_2^N - q_1^N)} \cdot \frac{f(k)}{a_2}. \qquad (40)$$

Notemos que la solución del problema de contorno (38) no existe solamente en el caso, cuando $q_1^N = q_2^N$, pero $q_1 \neq q_2$.

Examinemos ahora un caso particular de empleo de la fórmula (40). Supongamos que se exige resolver el primer problema de contorno para la ecuación

$$y(n+1) - 2xy(n) + y(n-1) = -f(n),$$
$$1 \leqslant n \leqslant N-1, \quad y(0) = \mu_1, \quad y(N) = \mu_2. \qquad (41)$$

Más arriba fueron halladas las raíces q_1 y q_2 de la ecuación característica correspondiente a (41)

$$q_1 = x + \sqrt{x^2 - 1}, \quad q_2 = x - \sqrt{x^2 - 1} = 1/q_1.$$

Sustituyendo estos valores en (40) y teniendo en cuenta la fórmula (17) para el polinomio $U_n(x)$, obtendremos la solución del problema (41), en la siguiente forma:

$$y(n)=\frac{U_{N-n-1}(x)}{U_{N-1}(x)}\left[\mu_1+\sum_{k=1}^{n-1}U_{k-1}(x)\,f(k)\right]+$$

$$+\frac{U_{n-1}(x)}{U_{N-1}(x)}\left[\mu_2+\sum_{k=n}^{N-1}U_{N-k-1}(x)\,f(k)\right]. \qquad (42)$$

La solución existe y se da por la fórmula (42), si se cumple la condición $x\neq\cos\frac{k\pi}{N}$, $k=1, 2, \ldots, N-1$. Regresemos a la ecuación (38). Si $a_0\,a_2>0$, entonces la solución (40) de este problema puede ser escrita en una forma más compacta que (40). En efecto, escribamos las raíces $q_1=\frac{1}{2a_2}[-a_1+\sqrt{a_1^2-4a_0a_2}]$, $q_2=\frac{1}{2a_2}[-a_1-\sqrt{a_1^2-4a_0a_2}]$ de la ecuación característica correspondiente a (38), en la siguiente forma:

$$q_1=\rho\,(x+\sqrt{x^2-1}),\qquad q_2=\rho\,(x-\sqrt{x^2-1}), \qquad (43)$$

donde

$$\rho=\sqrt{\frac{a_0}{a_2}},\qquad x=-\frac{a_1}{2\sqrt{a_0a_2}}. \qquad (44)$$

Sustituyamos (43) en (40) y tengamos en cuenta la fórmula (17). Obtendremos la solución del problema (38) para el caso $a_0a_2>0$ en la forma

$$y(n)=\frac{U_{N-n-1}(x)}{U_{N-1}(x)}\,\rho^n\left[\mu_1+\sum_{k=1}^{n-1}\frac{U_{k-1}(x)}{\rho^{k-1}}\cdot\frac{f(k)}{a_0}\right]+$$

$$+\frac{U_{n-1}(x)}{U_{N-1}(x)}\cdot\frac{1}{\rho^{N-n}}\left[\mu_2+\sum_{k=n}^{N-1}\rho^{N-k-1}U_{N-k-1}(x)\,\frac{f(k)}{a_0}\right],$$

donde ρ y x están definidas en (44). La solución del problema (38) para el caso $a_0a_2>0$ existe, si se cumple la condición

$$a_1+2\sqrt{a_0a_2}\cos\frac{k\pi}{N}\neq 0,\quad k=1, 2, \ldots, N-1.$$

Examinemos ahora el primer problema de contorno para la ecuación vectorial tripuntual con coeficientes constantes

$$\begin{aligned} &Y_{n-1} - CY_n + Y_{n+1} = -F_n, \quad 1 \leqslant n \leqslant N-1 \\ &Y_0 = F_0, \quad Y_N = F_N, \end{aligned} \tag{45}$$

donde Y_n y F_n son los vectores, y C es la matriz cuadrada. Es fácil comprobar, que la solución general de la ecuación no homogénea (45) tiene la forma

$$Y_n = U_{n-2}\left(\frac{1}{2}C\right)C_1 + U_{n-1}\left(\frac{1}{2}C\right)C_2 - \\ - \sum_{k=1}^{n-1} U_{n-k-1}\left(\frac{1}{2}C\right)F_k,$$

donde C_1 y C_2 son los vectores arbitrarios, y $U_n(X)$ es el polinomio matricial de la matriz X definido por las fórmulas recurrentes (19).

Si la matriz C es tal, que $U_{N-1}\left(\frac{1}{2}C\right)$ es la matriz no degenerada, entonces la solución del problema de contorno (45) se determina por una fórmula, análoga a la fórmula (42).

$$Y_n = U_{N-1}^{-1}\left(\frac{1}{2}C\right)U_{N-n-1}\left(\frac{1}{2}C\right) \times \\ \times \left[F_0 + \sum_{k=1}^{n-1} U_{k-1}\left(\frac{1}{2}C\right)F_k\right] + \\ + U_{N-1}^{-1}\left(\frac{1}{2}C\right)U_{n-1}\left(\frac{1}{2}C\right)\left[F_N + \sum_{k=n}^{N-1} U_{N-k-1}\left(\frac{1}{2}C\right)F_k\right]. \tag{46}$$

Más abajo será mostrado que el problema de Dirichlet de diferencias para la ecuación de Poisson en un rectángulo se reduce al problema (45).

Como conclusión observemos que a la condición de existencia de la solución del problema (45) se le puede dar la siguiente formulación: la solución existe y se determina por la fórmula (46), si los números $\cos\frac{k\pi}{N}$, $k = 1, 2, \ldots$ $\ldots, N-1$ no son valores propios de la matriz C.

§ 5. Problemas de diferencias sobre valores propios

1. Primer problema de contorno en valores propios. En el capítulo IV será examinado el método de separación de variables que se utiliza para encontrar las soluciones de los problemas de contorno reticulares para ecuaciones elípticas en un rectángulo. En relación con esto surge la necesidad de representar las funciones reticulares buscadas en forma de un desarrollo por las funciones propias del respectivo problema de diferencias. En este párrafo examinaremos problemas de diferencias sobre valores propios para el operador de diferencias de segundo orden más simple, definido en una red uniforme.

Formulemos el primer problema de contorno. Supongamos que está introducida una red uniforme $\bar{\omega} = \{x_i = ih, i = 0, 1, \ldots, N, hN = l\}$ con paso h, en el segmento $[0, l]$. Se exige hallar aquellos valores del parámetro λ (valores propios), para los cuales existen soluciones no triviales $\mu(x_i)$ (funciones propias) del siguiente problema de diferencias:

$$y_{\bar{x}x} + \lambda y = 0, \quad x \in \omega, \quad y(0) = y(l) = 0, \tag{1}$$

donde

$$y_{\bar{x}x,\, i} = \frac{y(i+1) - 2y(i) + y(i-1)}{h^2}, \quad y(i) = y(x_i).$$

Hallemos la solución del problema (1). Para eso escribamos (1) en forma del problema de contorno para la ecuación de diferencias de segundo orden

$$y(i+1) - 2\left(1 - \frac{h^2\lambda}{2}\right) y(i) + y(i-1) = 0, \quad 1 \leqslant i \leqslant N-1,$$
$$y(0) = y(N) = 0. \tag{2}$$

En el punto 1 del § 4 se mostró que la solución general de la ecuación (2) tiene la forma (véase la fórmula (20) del § 4) $y(i) = c_1 T_i(z) + c_2 U_{i-1}(z)$, donde c_1 y c_2 son las constantes arbitrarias, y mediante z aquí está denotado

$$z = 1 - h^2\lambda/2. \tag{3}$$

Determinaremos las constantes c_1 y c_2 de las condiciones de contorno

$$y(0) = c_1 = 0, \quad y(N) = c_2 U_{N-1}(z) = 0. \tag{4}$$

Aquí y a continuación aplicaremos las fórmulas (15) y (18) del § 4, que definen los polinomios de Chebishev de primero y segundo género y también las fórmulas (21)-(32) del mismo párrafo.

Ya que se busca la solución no trivial del problema (1), entonces $c_2 \neq 0$ y de (4) tendremos la condición $U_{N-1}(z) = 0$, al cumplir la cual la solución del problema (1) tiene la forma $y_i = c_2 U_{i-1}(z)$.

Puesto que los números $z_k = \cos \frac{k\pi}{N}$, $k = 1, 2, \ldots, N-1$ son las raíces del polinomio $U_{n-1}(z)$, entonces de (3) encontramos los valores propios del problema (1).

$$\lambda_k = \frac{4}{h^2} \operatorname{sen}^2 \frac{k\pi}{2N} = \frac{4}{h^2} \operatorname{sen}^2 \frac{k\pi h}{2l}, \quad k = 1, 2, \ldots, N-1. \tag{5}$$

A cada valor propio λ_k le corresponde la solución no nula del problema (1)

$$y_k(i) = c_2 U_{i-1}(z_k) = \bar{c}_k \operatorname{sen} \frac{k\pi i}{N} = \bar{c}_k \operatorname{sen} \frac{k\pi x_i}{l},$$

$$0 \leqslant i \leqslant N \quad \left(c_2 = \bar{c}_k \operatorname{sen} \frac{k\pi}{N}\right). \tag{6}$$

Definamos el producto escalar de funciones reticulares prefijadas sobre $\bar{\omega}$ de la siguiente forma:

$$(u, v) = \sum_{i=1}^{N-1} u(i) v(i) h + 0{,}5h [u(0) v(0) + u(N) v(N)].$$

Determinemos ahora la constante $\bar{c}_k$ en (6) de manera tal, que las funciones $y_k(i)$ tengan norma, igual a la unidad, es decir $(y_k, y_k) = 1$.

Cálculos no complicados dan $\bar{c}_k = \sqrt{2/l}$. Sustituyendo el valor hallado para $\bar{c}_k$ en (6), obtendremos las funciones propias $\mu_k(i)$ del problema (1)

$$\mu_k(i) = \sqrt{\frac{2}{l}} \operatorname{sen} \frac{k\pi i}{N} = \sqrt{\frac{2}{l}} \operatorname{sen} \frac{k\pi x_i}{l}, \tag{7}$$

$$i = 0, 1, \ldots, N, \quad k = 1, 2, \ldots, N-1.$$

Así está resuelto el problema (1) y la solución está dada en (5) y (7).

Enumeremos las propiedades fundamentales de las funciones propias y los valores propios del primer problema de contorno (1).

1) Las funciones propias están ortonormalizadas:

$$(\mu_k, \mu_m) = \delta_{km}, \quad \delta_{km} = \begin{cases} 1, & k = m, \\ 0, & k \neq m. \end{cases}$$

2) Para toda función reticular $f(i)$ prefijada en los nodos interiores de la red $\overline{\omega}$, es decir, para $1 \leqslant i \leqslant N - 1$, tiene lugar la descomposición

$$f(i) = \frac{2}{N} \sum_{k=1}^{N-1} \varphi_k \operatorname{sen} \frac{k\pi i}{N}, \quad i = 1, 2, \ldots, N-1. \qquad (8)$$

donde

$$\varphi_k = \sum_{i=1}^{N-1} f(i) \operatorname{sen} \frac{k\pi i}{N}, \quad k = 1, 2, \ldots, N-1. \qquad (9)$$

Expliquemos esta afirmación. Sea $f(i)$ una función reticular arbitraria definida sobre ω (o sobre $\overline{\omega}$ y que se anula para $i = 0$ e $i = N$). Desarrollémosla por las funciones propias

$$f(i) = \sum_{k=1}^{N-1} f_k \mu_k(i) = \sum_{k=1}^{N-1} \sqrt{\frac{2}{l}} f_k \operatorname{sen} \frac{k\pi i}{N}, \qquad (10)$$

donde f_k son los coeficientes de Fourier de la función $f(i)$. Multiplicando (10) escalarmente por $\mu_m(i)$ y utilizando la ortonormalidad de las funciones propias, hallamos los coeficientes de Fourier

$$f_m = \sum_{k=1}^{N-1} f_k (\mu_k, \mu_m) = (f, \mu_m) = \sum_{i=1}^{N-1} \sqrt{\frac{2}{l}} f(i) \operatorname{sen} \frac{\pi m i}{N} h.$$

La relación de las fórmulas obtenidas con (8)-(9) se establece fácilmente si observamos, que $f_m = \frac{\sqrt{2l}}{N} \varphi_m$.

El desarrollo (8)-(9) es cómodo porque para calcular la transformada de Fourier de la función $f(i)$ y para el restablecimiento de la función inicial por su transformada hay que calcular una suma de un solo tipo. En el capítulo IV será examinado un algoritmo para el cálculo rápido de sumas de tal tipo.

3. Para los valores propios son válidas las desigualdades

$$\frac{8}{l^2} \leqslant \frac{4}{h^2} \operatorname{sen}^2 \frac{\pi}{2N} = \lambda_1 \leqslant \lambda_k \leqslant \lambda_{N-1} = \\ = \frac{4}{h^2} \cos^2 \frac{\pi}{2N}, \quad 1 \leqslant k \leqslant N-1.$$

2. Segundo problema de contorno. Examinemos ahora el segundo problema de contorno en valores propios

$$\begin{gathered} y_{\bar{x}x} + \lambda y = 0, \quad x \in \omega, \\ \frac{2}{h} y_x + \lambda y = 0, \ x = 0, \ -\frac{2}{h} y_{\bar{x}} + \lambda y = 0, \ x = l. \end{gathered} \tag{11}$$

Hallemos la solución del problema (11). Anotando las derivadas de diferencias en (11) por puntos, obtendremos el problema

$$\begin{gathered} y(i+1) - 2zy(i) + y(i-1) = 0, \quad 1 \leqslant i \leqslant N-1, \\ y(1) - zy(0) = 0, \ y(N-1) - zy(N) = 0, \end{gathered} \tag{12}$$

donde $z = 1 - \lambda h^2/2$. De la solución general $y(i) = c_1 T_i(z) + c_2 U_{i-1}(z)$ de la ecuación (12), separemos la solución que satisface las condiciones de contorno planteadas. Aplicando la fórmula (24) del § 4, tendremos

$$y(1) - zy(0) = c_1 z + c_2 - c_1 z = c_2 = 0, \quad c_2 = 0,$$

y también

$$y(N-1) - zy(N) = c_1 (T_{N-1}(z) - zT_N(z)) = \\ = c_1 (1 - z^2) U_{N-1}(z) = 0.$$

Ya que $c_1 \neq 0$, de aquí obtenemos

$$z_k = \cos\frac{k\pi}{N}, \quad k = 0, 1, \ldots, N,$$

y, por consiguiente, los valores propios del problema (12) son

$$\lambda_k = \frac{4}{h^2} \operatorname{sen}^2 \frac{k\pi}{2N} = \frac{4}{h^2} \operatorname{sen}^2 \frac{k\pi h}{2l}, \quad k = 0, 1, \ldots, N. \tag{13}$$

A su vez, a cada λ_k le corresponde la solución no nula del problema (11)

$$y_k(i) = c_k T_i(z_k) = c_k \cos\frac{k\pi i}{N}, \quad 0 \leqslant i \leqslant N.$$

Elijamos las constantes c_k de la condición $(y_k, y_k) = 1$, donde el producto escalar está definido más arriba. Cálculos

directos muestran, que

$$c_k=\sqrt{2/l},\ k=1,\ 2,\ \ldots,\ N-1,\ c_k=\sqrt{1/l},\ k=0,\ N.$$

De esta forma, las funciones propias normadas del problema (11) son las funciones

$$\mu_k(i)=\sqrt{\frac{2}{l}}\cos\frac{k\pi i}{N}=\sqrt{\frac{2}{l}}\cos\frac{k\pi x_i}{l},\quad 1\leqslant k\leqslant N-1,$$
$$\mu_k(i)=\sqrt{\frac{1}{l}}\cos\frac{k\pi i}{N}=\sqrt{\frac{1}{l}}\cos\frac{k\pi x_i}{l},\quad k=0,\ N, \tag{14}$$

definidas sobre la red $\bar{\omega}$. Notemos que la función propia correspondiente al valor propio nulo $\lambda_0=0$, es la constante $\mu_0(i)=\sqrt{1/l}$.

Formulemos las propiedades de las funciones propias y los valores propios del segundo problema de contorno (11).

1) Las funciones propias están ortonormalizadas: $(\mu_k, \mu_m)=\delta_{km}$.

2) Para cualquier función reticular $f(i)$ definida sobre $\bar{\omega}$, tiene lugar el desarrollo

$$f(i)=\frac{2}{N}\sum_{k=0}^{N}\rho_k\varphi_k\cos\frac{k\pi i}{N},\quad i=0,\ 1,\ \ldots\ N, \tag{15}$$

donde

$$\varphi_k=\sum_{i=0}^{N}\rho_i f(i)\cos\frac{k\pi i}{N},\quad k=0,\ 1,\ \ldots,\ N, \tag{16}$$

$$\rho_i=\begin{cases}1, & 1\leqslant i\leqslant N-1,\\ 0,5, & i=0,\ N.\end{cases} \tag{17}$$

Las fórmulas (15) y (16) son una modificación del desarrollo tradicional de $f(i)$ por las funciones propias $\mu_k(i)$

$$f(i)=\sum_{k=0}^{N}f_k\mu_k(i),\quad f_k=(f,\ \mu_k)$$

mediante la siguiente sustitución:

$$f_k=\begin{cases}\dfrac{\sqrt{2l}}{N}\varphi_k, & 1\leqslant k\leqslant N-1,\\ \dfrac{1}{N}\sqrt{l}\varphi_k, & k=0,\ N.\end{cases}$$

3) Para los valores propios son válidas las desigualdades

$$0=\lambda_0 \leqslant \lambda_k \leqslant \lambda_N, \quad 0 \leqslant k \leqslant N.$$

3. Problema de contorno mixto. Examinemos ahora el problema en valores propios cuando en un lado del segmento $[0, l]$ está dada la condición de contorno de primer género y en el otro, la de segundo género, por ejemplo:

$$\begin{aligned} &y_{\bar{x}x}+\lambda y=0, \quad x \in \omega, \\ &y(0)=0, \qquad -\frac{2}{h} y_{\bar{x}}+\lambda y=0, \quad x=l. \end{aligned} \tag{18}$$

Tal problema lo llamaremos *problema de contorno mixto.*

Hallemos la solución del problema (18). El problema correspondiente a (18) para la ecuación de diferencias de segundo orden tiene la forma

$$\begin{aligned} &y(i+1)-2zy(i)+y(i-1)=0, \ 1 \leqslant i \leqslant N-1, \\ &y(0)=0, \quad y(N-1)-zy(N)=0, \end{aligned}$$

donde $z=1-0{,}5\lambda h^2$. De la solución general de esta ecuación

$$y(i)=c_1 T_i(z)+c_2 U_{i-1}(z)$$

separemos aquella solución que satisface las condiciones de contorno prefijadas. Utilizando (25) del § 4 obtendremos

$$\begin{aligned} &y(0)=c_1=0, \\ &y(N-1)-zy(N)=c_2(U_{N-2}(z)-zU_{N-1}(z))= \\ &\qquad\qquad\qquad\qquad =-c_2 T_N(z)=0. \end{aligned}$$

Puesto que $c_2 \neq 0$, de aquí hallamos $T_N(z_k)=0$, donde $z_k=\cos\frac{(2k-1)\pi}{2N}$, $k=1, 2, \ldots, N$ y, por lo tanto, los valores propios del problema (18) son los números

$$\lambda_k=\frac{4}{h^2}\operatorname{sen}^2\frac{(2k-1)\pi}{4N}=\frac{4}{h^2}\operatorname{sen}^2\frac{(2k-1)\pi h}{4l}, \tag{19}$$
$$k=1, 2, \ldots, N.$$

Las funciones propias normalizadas del problema (18), correspondientes a los valores propios λ_k, son

$$\begin{aligned} \mu_k(i)&=\sqrt{\frac{2}{l}}\operatorname{sen}\frac{(2k-1)\pi i}{2N}= \\ &=\sqrt{\frac{2}{l}}\operatorname{sen}\frac{(2k-1)\pi x_i}{2l}, \quad k=1, 2, \ldots, N, \end{aligned} \tag{20}$$

Formulemos las propiedades de las funciones propias y los valores propios del problema de contorno mixto (18).

1) Las funciones propias están ortonormalizadas: $(\mu_k, \mu_m) = \delta_{km}$.

2) Para toda función reticular $f(i)$ definida sobre $\omega^+ = \{x_i = ih, 1 \leqslant i \leqslant N\}$ (o sobre $\bar{\omega}$ y que se anula para $i = 0$) es válido el desarrollo

$$f(i) = \frac{2}{N} \sum_{k=1}^{N} \varphi_k \operatorname{sen} \frac{(2k-1)\pi i}{2N}, \quad i = 1, 2, \ldots, N, \tag{21}$$

donde

$$\varphi_k = \sum_{i=1}^{N} \rho_i f(i) \operatorname{sen} \frac{(2k-1)\pi i}{2N}, \quad k = 1, 2, \ldots, N, \tag{22}$$

y ρ_i está definido en (17).

3) Para los valores propios son válidas las desigualdades

$$\frac{8}{(2+\sqrt{2})\, l^2} \leqslant \frac{4}{h^2} \operatorname{sen}^2 \frac{\pi}{2N} =$$

$$= \lambda_1 \leqslant \lambda_k \leqslant \lambda_N = \frac{4}{h^2} \cos^2 \frac{\pi}{4N}, \quad 1 \leqslant k \leqslant N.$$

Si para la ecuación (18) está prefijada la condición de contorno de primer género en el extremo derecho del segmento $[0, l]$, es decir, está dado el problema

$$\begin{gathered} y_{\bar{x}x} + \lambda y = 0, \quad x \in \omega, \\ \frac{2}{h} y_x + \lambda y = 0, \quad x = 0; \quad y(l) = 0, \end{gathered} \tag{23}$$

entonces los valores propios se determinan por la fórmula (19), y las funciones propias normalizadas son

$$\mu_k(i) = \sqrt{\frac{2}{l}} \operatorname{sen} \frac{(2k-1)(N-i)\pi}{2N} =$$

$$= \sqrt{\frac{2}{l}} \operatorname{sen} \frac{(2k-1)\pi(l-x_i)}{2l}, \quad k = 1, 2 \ldots, N.$$

Tiene lugar la siguiente afirmación. Para *toda función reticular $f(i)$, prefijada sobre $\omega^- = \{x_i = ih, i = 0, 1, \ldots, N-1, hN = l\}$ (o sobre $\bar{\omega}$ y que se anula siendo*

$i = N$), *es válido el desarrollo*

$$f(N-i) = \frac{2}{N} \sum_{k=1}^{N} \varphi_k \operatorname{sen} \frac{(2k-1)\pi i}{2N},$$
$$i = 1, 2, \ldots, N, \qquad (24)$$

donde

$$\varphi_k = \sum_{i=1}^{N} \rho_{N-i} f(N-i) \operatorname{sen} \frac{(2k-1)\pi i}{2N},$$
$$k = 1, 2, \ldots, N, \qquad (25)$$

y ρ_i *está definido en* (17).

Notemos que las funciones propias construidas del problema (23) están también ortonormalizadas:

$$(\mu_k, \mu_m) = \delta_{km}.$$

4. **Problema de contorno periódico.** Supongamos que en la red $\Omega = \{x_i = ih, i = 0, \pm 1, \pm 2, \ldots\}$, introducida sobre la recta $-\infty < x < \infty$, se busca la solución periódica no trivial con período N del siguiente problema en valores propios:

$$y_{\bar{x}x} + \lambda y = 0, \quad x \in \Omega,$$
$$y(i+N) = y(i), \quad i = 0, \pm 1, \pm 2, \ldots, h = l/N. \qquad (26)$$

Como la solución es periódica, entonces es suficiente hallarla para $i = 0, 1, \ldots, N-1$. Escribiendo (26) por los puntos $i = 0, 1, \ldots, N-1$ y teniendo en cuenta que $y(-1) = y(N-1)$, $y(0) = y(N)$ obtendremos el siguiente problema:

$$y(i+1) - 2zy(i) + y(i-1) = 0, \quad 0 \leqslant i \leqslant N-1,$$
$$y(0) = y(N), \quad y(-1) = y(N-1), \qquad (27)$$

donde $z = 1 - 0{,}5\lambda h^2$.

Hallemos la solución del problema (27). Sustituyamos la solución general

$$y(i) = c_1 T_i(z) + c_2 U_{i-1}(z)$$

en las condiciones de contorno. Teniendo en cuenta las propiedades de los polinomios de Chebishov obtenemos el

siguiente sistema para determinar las constantes c_1 y c_2:

$$c_1 (1 - T_N (z)) - c_2 U_{N-1} (z) = 0,$$
$$c_1 (T_{N-1} (z) - z) + c_2 (1 + U_{N-2} (z)) = 0. \qquad (28)$$

Este sistema tiene solución no nula entonces y solamente entonces, cuando su determinante es igual a cero. Calculemos este determinante, aplicando para las transformaciones las fórmulas (25), (29) y (31) del § 4. Obtenemos

$$(1 - T_N (z)) (1 + U_{N-2} (z)) + (T_{N-1} (z) - z) U_{N-1} (z) =$$
$$= 1 + U_{N-2} (z) - zU_{N-1} (z) - T_N (z) + T_{N-1} (z) U_{N-1} (z) -$$
$$- T_N (z) U_{N-2} (z) = 2 [1 - T_N (z)] = 0.$$

De aquí se deduce, que cuando $z = z_k$, donde

$$z_k = \cos \frac{2k\pi}{N}, \quad k = 0, 1, \ldots, N-1, \qquad (29)$$

el sistema (28) tiene solución no nula. De esta forma, los valores propios del problema (26) son

$$\lambda_k = \frac{4}{h^2} \operatorname{sen}^2 \frac{k\pi}{N} = \frac{4}{h^2} \operatorname{sen}^2 \frac{k\pi h}{l},$$
$$k = 0, 1, \ldots, N-1. \qquad (30)$$

Obtengamos ahora la solución del sistema (28). De tal modo tienen lugar las igualdades

$$T_{N-1} (z_k) = z_k, \quad 0 \leqslant k \leqslant N-1,$$

$$U_{N-2} (z_k) = \begin{cases} N-1, & k = 0, N/2, \\ -1, & k \neq 0, N/2, \end{cases}$$

$$U_{N-1}(z_k) = \begin{cases} N, & k = 0, \\ -N, & k = N/2, \\ 0, & k \neq 0, N/2, \end{cases}$$

entonces, sustituyendo (29) en (28), hallaremos la siguiente solución del sistema (28):

a) para $k = 0$ y $k = N/2$ tenemos $c_2 = 0$ y $c_1 = c_1^{(k)} \neq \neq 0$;

b) para $k \neq 0$, $k \neq N/2$, $0 < k \leqslant N-1$, las constantes $c_1 = c_1^{(k)}$ y $c_2 = c_2^{(k)}$ son arbitrarias pero no son iguales a cero simultáneamente. De aquí obtenemos que las fun-

ciones

$$y_k(i) = c_1^{(k)} \cos \frac{2k\pi i}{N}, \quad k = 0,\ N/2,$$

$$y_k(i) = c_1^{(k)} \cos \frac{2k\pi i}{N} + c_2^{(k)} \operatorname{sen} \frac{2k\pi i}{N}, \tag{31}$$

$$1 \leqslant k \leqslant N-1, \quad k \neq 0,\ \frac{N}{2}$$

son las soluciones del problema (27), que corresponden al valor propio λ_k. Notemos que en el caso $k \neq 0$, $N/2$ las fórmulas (31) determinan en realidad dos funciones linealmente independientes $c_1^{(k)} \cos \frac{2k\pi i}{N}$ y $c_2^{(k)} \operatorname{sen} \frac{2k\pi i}{N}$, cada una de las cuales es la solución del problema (27) y corresponde al valor propio λ_k.

Construyamos ahora las funciones propias normadas del problema (26). Observemos que para las funciones reticulares periódicas el producto escalar introducido antes, se puede escribir de la siguiente forma:

$$(u, v)_{\bar\omega} = \sum_{i=1}^{N-1} u(i)\, v(i)\, h + 0{,}5h\, [u(0)\, v(0) + u(N)\, v(N)] = \sum_{i=0}^{N-1} u(i)\, v(i)\, h.$$

Examinemos dos casos. Sea primeramente N par. De (31) obtenemos que las funciones propias correspondientes a λ_0 y $\lambda_{N/2}$ son

$$\mu_k(i) = \sqrt{\frac{1}{l}} \cos \frac{2k\pi i}{N}, \quad k = 0,\ \frac{N}{2}. \tag{32}$$

A continuación notemos que de (30) se deducen las igualdades

$$\lambda_{N-k} = \frac{4}{h^2} \operatorname{sen}^2 \frac{(N-k)\pi}{N} = \frac{4}{h^2} \operatorname{sen}^2 \frac{k\pi}{N} = \lambda_k,$$

$$k = 1,\ 2,\ \ldots,\ \frac{N}{2} - 1.$$

Eligiendo en calidad de función propia correspondiente al valor propio λ_k, la función

$$\mu_k(i) = \sqrt{\frac{2}{l}} \cos \frac{2k\pi i}{N}, \quad 1 \leqslant k \leqslant \frac{N}{2} - 1$$

y la función

$$\mu_{N-k}(i) = \sqrt{\frac{2}{l}} \operatorname{sen} \frac{2k\pi i}{N}, \quad 1 \leqslant k \leqslant \frac{N}{2} - 1,$$

correspondiente al valor propio $\lambda_{N-k} = \lambda_k$, obtendremos junto con (32) un sistema completo de funciones propias del problema (26). Así, los valores propios son los λ_k definidos en (30), y las funciones propias del problema (26) se dan por las fórmulas

$$\mu_k(i) = \sqrt{\frac{1}{l}} \cos \frac{2k\pi i}{N}, \quad k = 0, \frac{N}{2},$$

$$\mu_k(i) = \sqrt{\frac{2}{l}} \cos \frac{2k\pi i}{N}, \quad 1 \leqslant k \leqslant \frac{N}{2} - 1, \tag{33}$$

$$\mu_k(i) = \sqrt{\frac{2}{l}} \operatorname{sen} \frac{2(N-k)\pi i}{N}, \quad \frac{N}{2} + 1 \leqslant k \leqslant N - 1$$

para el caso de N par.

Observemos las propiedades fundamentales de las funciones propias y valores propios del problema de contorno periódico (26).

1) Las funciones propias están ortonormalizadas.

2) Toda función reticular $f(i)$ periódica con período N definida sobre la red Ω, puede ser representada en la forma

$$f(i) = \frac{2}{N} \sum_{k=0}^{N/2} \rho_k \varphi_k \cos \frac{2k\pi i}{N} + \frac{2}{N} \sum_{k=N/2+1}^{N-1} \varphi_k \operatorname{sen} \frac{2(N-k)\pi i}{N}, \tag{34}$$

donde

$$\varphi_k = \sum_{i=0}^{N-1} \rho_k f(i) \cos \frac{2k\pi i}{N}, \quad 0 \leqslant k \leqslant \frac{N}{2},$$

$$\varphi_k = \sum_{i=0}^{N-1} f(i) \operatorname{sen} \frac{2(N-k)\pi i}{N}, \quad \frac{N}{2} + 1 \leqslant k \leqslant N - 1, \tag{35}$$

$$\rho_k = \begin{cases} 1, & k \neq 0, N/2 \\ 1/\sqrt{2}, & k = 0, N/2. \end{cases} \tag{36}$$

Las fórmulas (34)—(36) se deducen del desarrollo de la función $f(i)$ por las funciones propias $\mu_k(i)$:

$$f(i) = \sum_{k=0}^{N-1} f_k \mu_k(i), \quad f_k = (f, \mu_k)$$

al efectuar la sustitución $f_k = \frac{\sqrt{2l}}{N} \varphi_k$.

3) Para los valores propios se cumplen las desigualdades

$$0=\lambda_0\leqslant\lambda_k\leqslant\lambda_{N/2}=\frac{4}{h^2},\quad 0\leqslant k\leqslant N-1.$$

Examinemos ahora el caso, cuando N es impar. En este caso los valores propios del problema (26) se determinan por las fórmulas (30), pero además $\lambda_0 = 0$ y tiene lugar la igualdad $\lambda_{N-k} = \lambda_k$, $k = 1, 2, \ldots, (N-1)/2$.

Las funciones propias correspondientes a los valores propios λ_k se determinan mediante las siguientes fórmulas:

$$\begin{aligned}
&\mu_0(i)=\sqrt{\frac{1}{l}},\quad k=0,\\
&\mu_k(i)=\sqrt{\frac{2}{l}}\cos\frac{2k\pi i}{N},\quad 1\leqslant k\leqslant\frac{N-1}{2},\\
&\mu_k(i)=\sqrt{\frac{2}{l}}\operatorname{sen}\frac{2(N-k)\pi i}{N},\quad \frac{N+1}{2}\leqslant k\leqslant N-1.
\end{aligned}\tag{37}$$

Las funciones propias (37) están ortonormalizadas, y los valores propios λ_k satisfacen las desigualdades $0=\lambda_0<\lambda_k<$ $<\lambda_{\frac{N-1}{2}}=\frac{4}{h^2}\cos^2\frac{\pi}{2N}$, $0<k<N-1$. Además, cualquier función reticular $f(i)$ periódica (con período N (N — impar), definida sobre la red Ω, es representable en la forma

$$f(i)=\frac{2}{N}\sum_{k=0}^{(N-1)/2}\rho_k\varphi_k\cos\frac{2k\pi i}{N}+\frac{2}{N}\sum_{k=(N+1)/2}^{N-1}\times$$

$$\times\,\varphi_k\operatorname{sen}\frac{2(N-k)\pi i}{N},$$

donde

$$\varphi_k=\sum_{i=0}^{N-1}\rho_k f(i)\cos\frac{2k\pi i}{N},\quad 0\leqslant k\leqslant\frac{N-1}{2},$$

$$\varphi_k=\sum_{i=0}^{N-1}f(i)\operatorname{sen}\frac{2(N-k)\pi i}{N},\quad \frac{N+1}{2}\leqslant k\leqslant N-1,$$

siendo ρ_k definido más arriba.

Capítulo II

Método de factorización

En este capítulo se estudian distintas variantes del método directo de solución de las ecuaciones reticulares, el método de factorización. Se examina la aplicación del método para la solución, tanto de ecuaciones escalares, como vectoriales.

En el § 1, se construye y se investiga el método de factorización para ecuaciones tripuntuales escalares. El § 2 está dedicado a distintas variantes del método de factorización, aquí son examinadas las factorizaciones por flujos, cíclica y no monótona. En el § 3 se examinan las factorizaciones monótona y no monótona para ecuaciones escalares pentapuntuales. En el § 4 son construidos los algoritmos de la factorización matricial para ecuaciones vectoriales bipuntuales y tripuntuales y el método de factorización ortogonal para ecuaciones bipuntuales.

§ 1. Método de factorización para ecuaciones tripuntuales

1. Algoritmo del método. En el capítulo I fueron expuestos los métodos de solución de ecuaciones de diferencias con coeficientes constantes. El presente capítulo está dedicado a la construcción de métodos directos de solución de problemas de contorno para ecuaciones de diferencias tripuntuales y pentapuntuales con coeficientes variables, y también para ecuaciones vectoriales tripuntuales. Aquí serán estudiadas distintas variantes del método de factorización que representa en sí mismo el método de eliminación de Gauss, aplicado a sistemas especiales de ecuaciones algebraicas lineales y que tiene en cuenta la estructura de banda de la matriz del sistema.

Comencemos a examinar el método de factorización a partir del caso de las ecuaciones escalares. Supongamos

que se exige hallar la solución del siguiente sistema de ecuaciones tripuntuales.

$$\begin{aligned} &c_0y_0 - b_0y_1 = f_0, \quad i = 0,\\ &-a_iy_{i-1} + c_iy_i - b_iy_{i+1} = f_i, \quad 1 \leqslant i \leqslant N - 1, \quad (1)\\ &-a_Ny_{N-1} + c_Ny_N = f_N, \quad i = N, \end{aligned}$$

o en forma vectorial

$$\mathcal{A}Y = F,$$

donde $Y = (y_0, y_1, \ldots, y_N)$ es el vector de las incógnitas, $F = (f_0, f_1, \ldots, f_N)$ es el vector de los miembros derechos, y $\mathcal{A}$ es la matriz cuadrada de $(N + 1)(N + 1)$

$$\mathcal{A} = \begin{Vmatrix} c_0 & -b_0 & 0 & 0\ldots & 0 & 0 & 0 & 0\\ -a_1 & c_1 & -b_1 & 0\ldots & 0 & 0 & 0 & 0\\ 0 & -a_2 & c_2 & -b_2\ldots & 0 & 0 & 0 & 0\\ . & . & . & \ldots. & . & . & . & .\\ 0 & 0 & 0 & 0\ldots & -a_{N-2} & c_{N-2} & -b_{N-2} & 0\\ 0 & 0 & 0 & 0\ldots & 0 & -a_{N-1} & c_{N-1} & -b_{N-1}\\ 0 & 0 & 0 & 0\ldots & 0 & 0 & -a_N & c_N \end{Vmatrix}$$

con coeficientes reales o complejos.

Los sistemas del tipo (1) aparecen en la aproximación tripuntual de problemas de contorno para ecuaciones diferenciales ordinarias de segundo orden con coeficientes constantes y variables, y tambén en la realización de los esquemas de diferencias para ecuaciones en derivadas parciales. En el último caso con frecuencia se exige resolver no solamente el problema (1), sino una serie de problemas con miembros derechos distintos, pero además el número de problemas en la serie puede ser igual a varias decenas y centenas para un número de incógnitas en cada problema $N \approx 100$. Por eso es necesario elaborar métodos económicos de resolución de problemas del tipo (1), para los cuales el número de operaciones sea proporcional al número de incógnitas. Un tal método para el sistema (1) es el *método de factorización.*

La posibilidad de construcción de un método económico se encierra en la particularidad del sistema (1). La matriz $\mathcal{A}$ correspondiente a (1) pertenece a la clase de matrices enrarecidas: de $(N + 1)^2$ elementos hay no más de $3N + 1$ elementos no nulos. Además, ella posee estructura de banda (es una matriz tridiagonal). Esta distribución regular

de los elementos no nulos de la matriz $\mathcal{A}$ permite obtener fórmulas de cálculo muy simples para calcular la solución.

Pasemos a la construcción del algoritmo para resolver el sistema (1). Recordemos la sucesión de operaciones que se realizan en el método de eliminación de Gauss. En el primer paso se elimina la incógnita y_0 de todas las ecuaciones del sistema (1) para $i = 1, 2, \ldots, N$ con ayuda de la primera ecuación de (1), después de las ecuaciones transformadas para $i = 2, 3, \ldots, N$ se elimina la incógnita y_1 mediante la ecuación correspondiente a $i = 1$, etc. Como resultado obtenemos una ecuación respecto a y_N; con esto termina el paso directo del método. En el paso inverso para $i = N - 1, N - 2, \ldots, 0$, se encuentra y_i mediante los $y_{i+1}, y_{i+2}, \ldots, y_N$ ya hallados y los miembros derechos transformados.

Siguiendo la idea del método de Gauss realicemos la exclusión de las incógnitas en (1). Introduzcamos notaciones, poniendo $\alpha_1 = b_0/c_0$, $\beta_1 = f_0/c_0$ y escribamos (1) en la siguiente forma:

$$\begin{aligned} &y_0 - \alpha_1 y_1 = \beta_1, && i = 0, \\ &-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = f_i, && 1 \leqslant i \leqslant N-1, \\ &-a_N y_{N-1} + c_N y_N = f_N, && i = N. \end{aligned} \qquad (1')$$

Tomemos las dos primeras ecuaciones del sistema (1').

$$y_0 - \alpha_1 y_1 = \beta_1, \qquad -a_1 y_0 + c_1 y_1 - b_1 y_2 = f_1.$$

Multipliquemos la primera ecuación por a_1 y sumémosla con la segunda. Tendremos $(c_1 - a_1\alpha_1)\, y_1 - b_1 y_2 = f_1 + a_1\beta_1$ o después de dividir por $c_1 - a_1\alpha_1$

$$y_1 - \alpha_2 y_2 = \beta_2, \quad \alpha_2 = \frac{b_1}{c_1 - \alpha_1 a_1}, \quad \beta_2 = \frac{f_1 + a_1\beta_1}{c_1 - \alpha_1 a_1}.$$

Todas las restantes ecuaciones del sistema (1') no contienen y_0, por eso el proceso de eliminación se termina en este primer paso. Debido a esto, obtenemos el nuevo sistema «abreviado»

$$\begin{aligned} &y_1 - \alpha_2 y_2 = \beta_2, && i = 1, \\ &-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = f_i, && 2 \leqslant i \leqslant N-1, \\ &-a_N y_{N-1} + c_N y_N = f_N, && i = N, \end{aligned} \qquad (3)$$

el cual no contiene la incógnita y_0 y posee una estructura análoga a (1'). Si este sistema será resuelto, entonces la incógnita y_0 se halla por la fórmula $y_0 = \alpha_1 y_1 + \beta_1$. Al

sistema (3) se le puede de nuevo aplicar el método descrito de eliminación de las incógnitas. En el segundo paso será eliminada la incógnita y_1, en el tercero y_2, y así sucesivamente. Como resultado del l-ésimo paso obtendremos el sistema para las incógnitas $y_l, y_{l+1}, \ldots, y_N$

$$\begin{aligned} y_l - \alpha_{l+1} y_{l+1} &= \beta_{l+1}, & i &= l, \\ -a_i y_{i-1} + c_i y_i - b_i y_{i+1} &= f_i, & l+1 &\leqslant i \leqslant N-1, \\ -a_N y_{N-1} + c_N y_N &= f_N, & i &= N \end{aligned} \qquad (4)$$

y las fórmulas para encontrar y_i con los números $i \leqslant l-1$

$$y_i = \alpha_{i+1} y_{i+1} + \beta_{i+1}, \quad i = l-1, l-2, \ldots, 0. \qquad (5)$$

Evidentemente, los coeficientes α_l y β_l se encuentran por las fórmulas

$$\alpha_{l+1} = \frac{b_l}{c_l - a_l \alpha_l}, \quad \beta_{l+1} = \frac{f_l + a_l \beta_l}{c_l - a_l \alpha_l},$$

$$l = 1, 2, \ldots, \quad \alpha_1 = \frac{b_0}{c_0}, \quad \beta_1 = \frac{f_0}{c_0}.$$

Suponiendo en (4) $l = N-1$, obtenemos el sistema para y_N e y_{N-1}

$$y_{N-1} - \alpha_N y_N = \beta_N, \quad -a_N y_{N-1} + c_N y_N = f_N,$$

del cual encontramos $y_N = \beta_{N+1}$, $y_{N-1} = \alpha_N y_N + \beta_N$.

Uniendo ambas igualdades con (5) ($l = N-1$) obtenemos las fórmulas finales para encontrar las incógnitas:

$$\begin{aligned} y_i &= \alpha_{i+1} y_{i+1} + \beta_{i+1}, \quad i = N-1, N-2, \ldots, 0, \\ y_N &= \beta_{N+1}, \end{aligned} \qquad (6)$$

donde α_i y β_i se obtienen mediante las fórmulas recurrentes

$$\alpha_{i+1} = \frac{b_i}{c_i - a_i \alpha_i}, \quad i = 1, 2, \ldots, N-1, \quad \alpha_1 = \frac{b_0}{c_0}, \qquad (7)$$

$$\beta_{i+1} = \frac{f_i + a_i \beta_i}{c_i - a_i \alpha_i}, \quad i = 1, 2, \ldots, N, \quad \beta_1 = \frac{f_0}{c_0}. \qquad (8)$$

Así, las fórmulas (6)-(8) describen el método de Gauss el cual en su aplicación al sistema (1) ha obtenido el nombre especial: *método de factorización*. Los coeficientes α_i y β_i se llaman *coeficientes de factorización*, las fórmulas (7) y (8) describen el *paso directo de la factorización*, y (6) el *paso inverso*. Puesto que los valores de y_i se encuentran aquí sucesivamente al pasar de $i+1$ a i, entonces las fórmulas (6)-(8) son algunas veces llamadas fórmulas de *factorización derecha*.

El cálculo elemental de las operaciones aritméticas en (6)-(8) muestra, que la realización del método de factorización por estas fórmulas exige el cumplimiento de $3N$ multiplicaciones, $2N + 1$ divisiones y $3N$ sumas y restas. Si no hacemos distinción entre las operaciones aritméticas, su número total para el método de factorización es $Q = 8N + 1$. De este número, $3N - 2$ se gastan en el cálculo de α_i y $5N + 3$ en el cálculo de β_i e y_i.

Notemos que los coeficientes α_i no dependen de la parte derecha del sistema (1), y se determinan solamente por los coeficientes a_i, b_i y c_i de las ecuaciones de diferencias. Por eso si se exige resolver la serie de problemas (1) con segundos miembros diferentes, pero con la misma matriz $\mathcal{A}$, los coeficientes de factorización α_i se calculan solamente al resolver el primer problema de la serie. Para cada problema ulterior se determinan sólo los coeficientes β_i y la solución y_i, pero además se emplean los α_i hallados antes. De esta forma, únicamente en la solución del primer problema de la serie se emplea el número $Q = 8N + 1$ de operaciones aritméticas, y en la solución de cada problema siguiente se gastarán ya tan sólo $5N + 3$ operaciones.

Como conclusión indiquemos el orden del cálculo según las fórmulas del método de factorización. A partir de α_1 y β_1 por las fórmulas (7) y (8) se determinan y se guardan en la memoria los coeficientes de factorización α_i y β_i. Después por las fórmulas (6) se encuentra la solución y_i.

2. **Método de las factorizaciones opuestas.** Más arriba fueron obtenidas las fórmulas de factorización derecha para resolver el sistema (1). Análogamente se deducen las fórmulas de *factorización izquierda*:

$$\xi_i = \frac{a_i}{c_i - b_i\xi_{i+1}}, \quad i = N-1,\ N-2,\ \ldots,\ 1,\ \xi_N = \frac{a_N}{c_N}, \tag{9}$$

$$\eta_i \frac{f_i + b_i\eta_{i+1}}{c_i - b_i\xi_{i+1}}, \quad i = N-1,\ N-2,\ \ldots,\ 0,\ \eta_N = \frac{f_N}{c_N}, \tag{10}$$

$$y_{i+1} = \xi_{i+1}y_i + \eta_{i+1}, \quad i = 0,\ 1,\ \ldots, N-1,\ y_0 = \eta_0. \tag{11}$$

Aquí los valores de y_i se encuentran sucesivamente al crecer el índice i (de izquierda a derecha).

A veces resulta cómodo combinar las factorizaciones derecha e izquierda, obteniéndose el así llamado método de *factorizaciones opuestas*. Este método es útil aplicarlo si hay necesidad de hallar solamente una incógnita, por ejemplo y_m $(0 \leqslant m \leqslant N)$ o un grupo de incógnitas consecutivas.

Obtongamos las fórmulas del método de factorizaciones opuestas. Sea $1 \leqslant m \leqslant N$ y supongamos que por las fórmulas (7)-(10) están halladas $\alpha_1, \alpha_2, \ldots, \alpha_m, \beta_1, \beta_2, \ldots \ldots, \beta_m$ y $\xi_N, \xi_{N-1}, \ldots, \xi_m, \eta_N, \eta_{N-1}, \ldots, \eta_m$. Escribamos las fórmulas (6) y (11) para el paso inverso de las factorizaciones derecha e izquierda para $i = m - 1$. Tendremos el sistema

$$y_{m-1} = \alpha_m y_m + \beta_m, \quad y_m = \xi_m y_{m-1} + \eta_m,$$

del cual hallamos y_m:

$$y_m = \frac{\eta_m + \xi_m \beta_m}{1 - \xi_m \alpha_m}.$$

Utilizando el y_m hallado, mediante las fórmulas (6) para $i = m - 1, m - 2, \ldots, 0$ hallaremos sucesivamente $y_{m-1}, y_{m-2}, \ldots, y_0$, y por las fórmulas (11) para $i = m, m + 1, \ldots, N$ calculamos los restantes $y_{m+1}, y_{m+2}, \ldots, y_N$.

De esta forma, las fórmulas del método de factorizaciones opuestas tienen la forma:

$$\begin{aligned}
&\alpha_{i+1} = \frac{b_i}{c_i - a_i\alpha_i}, \quad i = 1, 2, \ldots, m-1, \quad \alpha_1 = \frac{b_0}{c_0},\\
&\beta_{i+1} = \frac{f_i + a_i\beta_i}{c_i - a_i\alpha_i}, \quad i = 1, 2, \ldots, m-1, \quad \beta_1 = \frac{f_0}{c_0},\\
&\xi_i = \frac{a_i}{c_i - b_i\xi_{i+1}}, \quad i = N-1, N-2, \ldots, m, \quad \xi_N = \frac{a_N}{c_N},\\
&\eta_i = \frac{f_i + b_i\eta_{i+1}}{c_i - b_i\xi_{i+1}}, \quad i = N-1, N-2, \ldots m, \quad \eta_N = \frac{f_N}{c_N}
\end{aligned} \tag{12}$$

para calcular los coeficientes de factorización y

$$\begin{aligned}
&y_i = \alpha_{i+1}y_{i+1} + \beta_{i+1}, \quad i = m-1, m-2, \ldots, 0,\\
&y_{i+1} = \xi_{i+1}y_i + \eta_{i+1}, \quad i = m, m+1, \ldots, N-1,\\
&y_m = \frac{\eta_m + \xi_m\beta_m}{1 - \xi_m\alpha_m}
\end{aligned} \tag{13}$$

para determinar la solución.

Es obvio, que el número de operaciones que se gastan para encontrar la solución del problema (1) mediante el método de factorizaciones opuestas, es el mismo que para la factorización izquierda o la derecha, es decir $Q \approx 8N$. Observemos que para el caso particular de coeficientes constantes $a_i = b_i = 1$, $c_i = c$ para $i = 1, 2, \ldots, N - 1$ y $b_0 = a_N = 0$ el número de operaciones puede ser dismi-

nuido, si N es un número impar, de la siguiente manera. Dado $N = 2M - 1$. Pongamos $m = M$ en las fórmulas (12) y (13) del método de factorizaciones opuestas. Entonces $\xi_{N-i+1} = \alpha_i$, $i = 1, 2, \ldots, M$. Por lo tanto, no es necesario hallar el coeficiente de factorización ξ_i y las fórmulas del método de factorizaciones opuestas tendrán la forma

$$\alpha_{i+1} = \frac{1}{c-\alpha_i}, \quad i = 1, 2, \ldots, M-1, \quad \alpha_1 = 0,$$

$$\beta_{i+1} = (f_i + \beta_i)\alpha_{i+1}, \quad i = 1, 2, \ldots, M-1, \quad \beta_1 = \frac{f_0}{c_0},$$

$$\eta_i = (f_i + \eta_{i+1})\alpha_{N-i+1}, \quad i = N-1, N-2, \ldots, M, \quad \eta_N = \frac{f_N}{c_N},$$

$$y_i = \alpha_{i+1}y_{i+1} + \beta_{i+1}, \quad i = M-1, M-2, \ldots, 0,$$

$$y_{i+1} = \alpha_{N-i}y_i + \eta_{i+1}, \quad i = M, M+1, \ldots, N-1,$$

donde $y_M = (\eta_M + \alpha_M\beta_M)/(1 - \alpha_M^2)$.

3. **Fundamentación del método de factorización.** Más arriba fueron obtenidas las fórmulas del método de factorización sin suposición alguna respecto a los coeficientes del sistema (1). Detengámonos aquí en la pregunta sobre cuáles exigencias deben satisfacer estos coeficientes, para que el método pueda ser aplicado con suficiente exactitud.

Aclaremos la situación. Ya que las fórmulas de cálculo (6)-(8) del método de factorización contienen operaciones de división, entonces es preciso garantizar que el denominador $c_i - a_i\alpha_i$ en (7) y (8) no se anula. Diremos que el algoritmo del método de factorización derecha *es correcto*, si $c_i - a_i\alpha_i \neq 0$ para $i = 1, 2, \ldots, N$. Más adelante la solución y_i se encuentra por la fórmula recurrente (6). Esta fórmula puede proporcionar acumulación de errores de redondeo de los resultados de las operaciones aritméticas. En efecto, supongamos que los coeficientes de factorización α_i y β_i son hallados exactamente, y al calcular y_N se admite un error ε_N, es decir, está hallado $\tilde{y}_N = y_N + \varepsilon_N$. Como la solución $\tilde{y}_i$ se encuentra mediante las fórmulas (6) $\tilde{y}_i = \alpha_{i+1}\tilde{y}_{i+1} + \beta_{i+1}$, $i = N-1, N-2, \ldots, 0$, entonces el error $\varepsilon_i = \tilde{y}_i - y_i$, obviamente satisfacerá la ecuación homogénea $\varepsilon_i = \alpha_{i+1}\varepsilon_{i+1}$, $i = N-1, N-2, \ldots, 0$ para ε_N dado. De aquí se deduce, que si todos los α_i son mayores en módulo que la unidad, entonces puede ocurrir un fuerte aumento del error ε_0 y, si N es suficientemente grande,

la solución real y_i obtenida se diferenciará significativamente de la solución y_i buscada.

Por no tener posibilidad de detenernos más detalladamente en la discusión de los problemas de estabilidad computacional del método y del mecanismo de formación de la inestabilidad, formularemos la exigencia que frecuentemente se plantea al algoritmo del método de factorización. Exigiremos que los coeficientes de factorización α_i no excedan en módulo la unidad. Esta condición suficiente garantiza el no crecimiento del error ε_i en la situación modelo examinada más arriba. Si se satisfaga la condición $|\alpha_i| \leqslant 1$, diremos que el algoritmo de factorización derecha *es estable*.

Aclaremos las condiciones de corrección y estabilidad del algoritmo (6)-(8). El siguiente lema contiene condiciones suficientes de corrección y estabilidad del algoritmo de factorización derecha.

LEMA 1. *Sean los coeficientes del sistema (1) reales y que satisfacen las condiciones* $|b_0| \geqslant 0$, $|a_N| \geqslant 0$, $|c_0| > 0$, $|c_N| > 0$, $|a_i| > 0$, $|b_i| > 0$, $i = 1, 2, \ldots, N-1$,

$$|c_i| \geqslant |a_i| + |b_i|, \quad i = 1, 2, \ldots, N-1, \tag{14}$$

$$|c_0| \geqslant |b_0|, \qquad |c_N| \geqslant |a_N|, \tag{15}$$

con todo al menos en una de las desigualdades (14) ó (15) se cumple la desigualdad estricta, es decir, la matriz $\mathcal{A}$ *posee una predominancia diagonal. Entonces para el algoritmo (6)-(8) del método de factorización tienen lugar las desigualdades* $c_i - a_i\alpha_i \neq 0$, $|\alpha_i| \leqslant 1$, $i = 1, 2, \ldots, N$ *que garantizan la corrección y estabilidad del método.*

Realicemos la demostración del lema por inducción. De las condiciones del lema y de (7) se desprende que

$$0 \leqslant |\alpha_1| = \frac{|b_0|}{|c_0|} \leqslant 1. \tag{16}$$

Mostremos, que de la desigualdad $|\alpha_i| \leqslant 1$ $(i \leqslant N-1)$ y de las condiciones del lema se deducen las desigualdades

$$c_i - a_i\alpha_i \neq 0, \quad |\alpha_{i+1}| \leqslant 1, \quad i \leqslant N-1. \tag{17}$$

Por tanto, teniendo en cuenta (16), obtendremos que tienen lugar las desigualdades $|\alpha_i| \leqslant 1$ para $i = 1, 2, \ldots, N$ y $c_i - a_i\alpha_i \neq 0$ para $i = 1, 2, \ldots, N-1$. Para culminar la demostración del lema queda por demostrar la desigualdad $c_N - a_N\alpha_N \neq 0$. Así, establezcamos primeramente (17).

Sea $|\alpha_i| \leqslant 1$, $i \leqslant N - 1$. Entonces de (14)

$$|c_i - a_i\alpha_i| \geqslant |c_i| - |a_i|\,|\alpha_i| \geqslant |b_i| + \\ + |a_i|(1 - |\alpha_i|) \geqslant |b_i| > 0, \qquad (18)$$

y, por consiguiente, $c_i - a_i\alpha_i \neq 0$. Más adelante, de (7) y (8) obtenemos

$$|\alpha_{i+1}| = \frac{|b_i|}{|c_i - a_i\alpha_i|} \leqslant \frac{|b_i|}{|b_i|} = 1,$$

lo que se exigía demostrar.

Queda por mostrar que $c_N - a_N\alpha_N \neq 0$. Para esto utilicemos la suposición de que al menos una de las desigualdades (14) ó (15) es estricta.

Aquí son posibles varios casos. Si $|c_N| > |a_N|$, entonces en virtud de lo demostrado $|\alpha_N| \leqslant 1$ y, por lo tanto, $c_N - a_N\alpha_N \neq 0$. Si la desigualdad estricta se alcanza en (14) para un cierto i_0, $1 \leqslant i_0 \leqslant N - 1$, entonces de (18) obtendremos, que $|c_{i_0} - a_{i_0}\alpha_{i_0}| > |b_{i_0}|$ y, por consiguiente, ocurre la desigualdad $|\alpha_{i_0+1}| < 1$. A continuación por inducción se establece fácilmente la desigualdad $|\alpha_i| < 1$ para $i \geqslant i_0 + 1$. Luego, en este caso tendremos $|\alpha_N| < 1$ y por eso $c_N - a_N\alpha_N \neq 0$. Si $|c_0| > |b_0|$, entonces la desigualdad $|\alpha_i| < 1$ tiene lugar comenzando desde $i = 1$. Por eso de nuevo obtenemos $|\alpha_N| < 1$ y $c_N - a_N\alpha_N \neq 0$. El lema está demostrado.

OBSERVACION 1. Las condiciones de corrección y estabilidad del algoritmo (6)-(8), formuladas en el lema 1, son sólo condiciones suficientes. Estas condiciones pueden ser debilitadas, permitiéndole a algunos de los coeficientes a_i y b_i anularse. Así por ejemplo, si para cierto $1 \leqslant m \leqslant N - 1$ resulta que $a_m = 0$, entonces el sistema (1) se descompone en dos sistemas:

$$\begin{aligned} &c_m y_m - b_m y_{m+1} = f_m, && i = m, \\ &-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = f_i, && m+1 \leqslant i \leqslant N - 1, \\ &-a_N y_{N-1} + c_N y_N = f_N, && i = N \end{aligned}$$

para las incógnitas $y_m, y_{m+1}, \ldots, y_N$ y

$$\begin{aligned} &c_0 y_0 - b_0 y_1 = f_0, && i = 0, \\ &-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = f_i, && 1 \leqslant i \leqslant m - 2, \\ &-a_{m-1} y_{m-2} + c_{m-1} y_{m-1} = f_{m-1} + b_{m-1} y_m \end{aligned}$$

para las incógnitas $y_0, y_1, \ldots, y_{m-1}$. A cada uno de estos sistemas se le puede aplicar el algoritmo (6)—(8), si para ellos se cumplen las condiciones del lema (1). Pero en este caso las fórmulas (6)—(8) pueden utilizarse para encontrar directamente la solución de todo el sistema (1) dividido y además el algoritmo será correcto y estable.

OBSERVACIÓN 2. Las condiciones del lema 1 garantizan la corrección y estabilidad de los algoritmos de las factorizaciones izquierda y opuestas. Estas condiciones se conservan también para el caso del sistema (1) con coeficientes complejos a_i, b_i y c_i.

Mostremos ahora, que cuando se cumplen las condiciones del lema 1 el sistema (1) tiene solución única para cualquier miembro derecho. En efecto, teniendo en cuenta las relaciones (7), se puede mostrar por una multiplicación directa de matrices que la matriz $\mathcal{A}$ del sistema (1) se representa en forma del producto de dos matrices triangulares L y U

$$\mathcal{A} = LU,$$

donde

$$L = \left\| \begin{matrix} c_0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ -a_1 & \Delta_1 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & -a_2 & \Delta_2 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & -a_3 & \Delta_3 & \ldots & 0 & 0 & 0 & 0 \\ . & . & . & . & \ldots & . & . & . & . \\ 0 & 0 & 0 & 0 & \ldots & \Delta_{N-3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & -a_{N-2} & \Delta_{N-2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & -a_{N-1} & \Delta_{N-1} & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & -a_N & \Delta_N \end{matrix} \right\|$$

$$U = \left\| \begin{matrix} 1 & -\alpha_1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -\alpha_2 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 1 & -\alpha_3 & \ldots & 0 & 0 & 0 \\ . & . & . & . & \ldots & . & . & . \\ 0 & 0 & 0 & 0 & \ldots & 1 & -\alpha_{N-1} & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 & -\alpha_N \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 \end{matrix} \right\|$$

y $\Delta_i = c_i - a_i\alpha_i$, $i = 1, 2, \ldots, N$. Ya que

$$\det \mathcal{A} = \det L \cdot \det U = c_0 \prod_{i=1}^{N} \Delta_i,$$

y en virtud del lema 1 $c_0 \neq 0$ y $\Delta_i = 0$ para $i = 1, 2, \ldots, N$, entonces det $\mathcal{A} \neq 0$. Por eso el sistema (1) en caso de cumplirse las condiciones del lema 1 posee solución única y esta solución puede ser hallada por el método de factorización (6)—(8).

4. Ejemplos de aplicación del método de factorización. Examinemos algunos ejemplos de aplicación del método de factorización expuesto más arriba.

EJEMPLO 1. *Primer problema de contorno.* Supongamos que se exige resolver el siguiente problema:

$$(k(x)\,u'(x))' - q(x)\,u(x) = -f(x), \quad 0 < x < l,$$
$$u(0) = \mu_1, \; u(l) = \mu_2, \; k(x) \geqslant c_1 > 0, \; q(x) \geqslant 0. \quad (19)$$

En el segmento $0 \leqslant x \leqslant l$ construyamos una red no uniforme arbitraria $\bar{\omega} = \{x_i \in [0, l], \; i = 0, 1, \ldots, N, \; x_0 = 0, \; x_N = l\}$ con pasos $h_i = x_i - x_{i-1}$, $i = 1, 2, \ldots, N$ y sustituyamos (19) por el siguiente problema de diferencias:

$$(ay_{\bar{x}})_{\hat{x}, i} - d_i y_i = -\varphi_i, \quad 1 \leqslant i \leqslant N-1,$$
$$y_0 = \mu_1, \; y_N = \mu_2, \quad (20)$$

donde $d_i = q(x_i)$, $\varphi_i = f(x_i)$, y para a_i utilicemos la aproximación más simple del coeficiente $k(x)$: $a_i = k(x_i - 0{,}5h_i)$. Anotando la derivada de diferencias que entra en (20) por puntos

$$(ay_{\bar{x}})_{\hat{x}, i} = \frac{1}{\hbar_i}\left(a_{i+1}\frac{y_{i+1} - y_i}{h_{i+1}} - a_i \frac{y_i - y_{i-1}}{h_i}\right),$$

donde $\hbar_i = 0{,}5(h_i + h_{i+1})$ es el paso medio en el punto x_i, obtendremos, que el problema (20) se escribe en forma del sistema

$$\begin{aligned} C_0 y_0 - B_0 y_1 &= f_0, & i &= 0, \\ -A_i y_{i-1} + C_i y_i - B_i y_{i+1} &= f_i, & 1 &\leqslant i \leqslant N-1, \\ -A_N y_{N-1} + C_N y_N &= f_N, & i &= N. \end{aligned} \quad (1'')$$

Aquí

$$B_0 = A_N = 0, \; C_0 = C_N = 1, \; f_0 = \mu_1, \; f_N = \mu_2, \; f_i = \varphi_i,$$
$$A_i = \frac{a_i}{\hbar_i h_i}, \; B_i = \frac{a_{i+1}}{\hbar_i h_{i+1}}, \; C_i = A_i + B_i + d_i, \quad 1 \leqslant i \leqslant N-1. \quad (21)$$

En virtud de la construcción del esquema de diferencias (20) para los coeficientes a_i y d_i se cumplen las siguientes

condiciones: $a_i \geqslant c_i > 0$, $d_i \geqslant 0$. Por eso de (21) se deduce que para (1″) están cumplidas las condiciones del lema 1 y este problema puede ser resuelto por el método de factorización.

EJEMPLO 2. *Tercer problema de contorno*. Examinemos ahora el caso de condiciones de contorno de tercer género:

$$\begin{aligned} &(k(x)\,u'(x))' - q(x)\,u(x) = -f(x), \qquad 0 < x < l, \\ &k(0)\,u'(0) = \varkappa_1 u(0) - \mu_1, \qquad\qquad (22) \\ &-k(l)\,u'(l) = \varkappa_2 u(l) - \mu_2. \end{aligned}$$

Consideraremos cumplidas las condiciones siguientes: $k(x) \geqslant c_1 > 0$, $q(x) \geqslant 0$, $\varkappa_1 \geqslant 0$, $\varkappa_2 \geqslant 0$, además, si $q(x) \equiv 0$, entonces $\varkappa_1^2 + \varkappa_2^2 \neq 0$.

En la red no uniforme introducida más arriba el problema (22) se aproxima por el siguiente esquema de diferencias:

$$\begin{aligned} &(ay_{\bar{x}})_{\hat{x},i} - d_i y_i = -\varphi_i, \qquad 1 \leqslant i \leqslant N-1, \\ &\frac{2}{h_1} a_1 y_{x,0} = \left(d_0 + \frac{2}{h_1}\varkappa_1\right) y_0 - \left(\varphi_0 + \frac{2}{h_1}\mu_1\right), \quad i = 0, \qquad (23) \\ &-\frac{2}{h_N} a_N y_{\bar{x},N} = \left(d_N + \frac{2}{h_N}\varkappa_2\right) y_N - \left(\varphi_N + \frac{2}{h_N}\mu_2\right), \quad i = N, \end{aligned}$$

donde los coeficientes a_i, d_i y φ_i son elegidos mediante el método indicado en el ejemplo 1. Anotando la segunda derivada de diferencias $(ay_{\bar{x}})_{\hat{x}}$ por puntos, y también las primeras derivadas

$$y_{x,i} = \frac{y_{i+1} - y_i}{h_{i+1}}, \quad y_{\bar{x},i} = \frac{y_i - y_{i-1}}{h_i},$$

reduciremos (23) a la forma (1″), donde

$$B_0 = \frac{2a_1}{h_1^2}, \quad A_N = \frac{2a_N}{h_N^2}, \quad C_0 = B_0 + d_0 + \frac{2}{h_1}\varkappa_1,$$

$$C_N = A_N + d_N + \frac{2}{h_N}\varkappa_2, \quad f_0 = \varphi_0 + \frac{2}{h_1}\mu_1, \quad f_N = \varphi_N + \frac{2}{h_N}\mu_2,$$

$$A_i = \frac{a_i}{h_i \hbar_i}, \quad B_i = \frac{a_{i+1}}{\hbar_i h_{i+1}}, \quad C_i = A_i + B_i + d_i, \quad f_i = \varphi_i,$$

$$1 \leqslant i \leqslant N-1.$$

Es fácil comprobar que en este caso también se cumplen las condiciones del lema 1.

EJEMPLO 3. *Esquemas de diferencias para la ecuación de conducción del calor*. Examinemos el primer problema

de contorno para la ecuación de conducción del calor:

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2},\quad 0<x\leqslant l,\ t>0,$$
$$u(0,\ t)=\mu_1(t),\qquad u(l,\ t)=\mu_2(t), \tag{24}$$
$$u(x,\ 0)=u_0(x).$$

En el plano $(x,\ t)$ introduzcamos la red $\overline{\omega}=\{(x_i,\ t_n),\ x_i=ih,\ i=0,\ 1,\ \ldots,\ N,\quad h=l/N,\quad t_n=n\tau,\ n=0,\ 1,\ \ldots\}$ con paso h en el espacio y τ según el tiempo. Aproximemos (24) por el esquema de diferencias

$$y_{t,\,i}=\sigma y^{n+1}_{\bar{x},\,x,\,i}+(1-\sigma)\,y^{n}_{\bar{x}x,\,i},\quad 1\leqslant i\leqslant N-1,$$
$$y^n_0=\mu_1(t_n),\ y^n_N=\mu_2(t_n),\ y^0_i=u_0(x_i),\ n=0,\ 1,\ \ldots, \tag{25}$$

dondo σ es el parámetro real, $y^n_i=y(x_i,\ t_n)$,

$$y_{\bar{x}x,\,i}=\frac{1}{h^2}(y_{i+1}-2y_i+y_{i-1}),$$
$$y_{t,\,i}=\frac{1}{\tau}(y^{n+1}_i-y^n_i). \tag{26}$$

Es conocido (véase, por ejemplo, [9]) que el esquema (25) tiene la aproximación $O(\tau+h^2)$ para toda σ; $O(\tau^2+h^2)$ para $\sigma=0{,}5$ y la aproximación $O(\tau^2+h^4)$ para $\sigma=1/2-h^2/(12\tau)$. La condición de estabilidad del esquema (25) según los datos iniciales tiene la forma

$$\sigma\geqslant 1/2-h^2/(4\tau). \tag{27}$$

Volvamos ahora al método de solución de las ecuaciones (25) respecto a y^{n+1}_i. Considerando y^n_i ya conocido, escribamos (25) en la siguiente forma:

$$\frac{1}{\sigma\tau}y^{n+1}_i-y^{n+1}_{\bar{x}x,\,i}=\varphi^n_i,\quad 1\leqslant i\leqslant N-1,$$
$$y^{n+1}_0=\mu_1(t_{n+1}),\ y^{n+1}_N=\mu_2(t_{n+1}),$$

donde $\varphi^n_i=\frac{1}{\sigma\tau}\ y^n_i+\left(\frac{1}{\sigma}-1\right)y^n_{\bar{x}x,\,i}$, si $\sigma\neq 0$.

Empleando (26) reduzcamos este esquema a la forma (1''), donde $B_0=A_N=0$, $c_0=C_N=1$, $f_0=\mu_1(t_{n+1})$, $f_N=\mu_2(t_{n+1})$, $A_i=B_i=\frac{1}{h^2}$, $C_i=A_i+B_i+\frac{1}{\sigma\tau}$, $f_i=\varphi^n_i$, $1\leqslant i\leqslant N-1$.

Hallemos las condiciones para las cuales se pueda resolver el sistema (1″) construido con ayuda del método de factorización. Del lema 1 se deduce, que se debe cumplir la condición $|2/h^2 + 1/(\sigma\tau)| \geqslant 2/h^2$. Resolviendo esta desigualdad, hallamos la condición suficiente de aplicabilidad de la factorización: $\sigma \geqslant -h^2/(4\tau)$. Comparando esta desigualdad con (27), obtendremos que si para el esquema (25) está cumplida la condición de estabilidad (27), entonces para encontrar la solución en la capa superior puede aplicarse el método de factorización.

EJEMPLO 4. *Ecuación de Shrodinguer no estacionaria.* Examinemos la ecuación de Shrodinguer no estacionaria $i\dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2}$, $0 < x < l$, $t > 0$, $u(0, t) = u(l, t) = 0$, $u(0, x) = u_0(x)$, $i = \sqrt{-1}$.

Para esta ecuación, al igual que para la ecuación de conducción del calor (24), se puede construir un esquema de dos capas con pesos

$$\begin{aligned} &iy_{t,k} = \sigma y^{n+1}_{\bar{x}x,k} + (1-\sigma)\, y^{n}_{\bar{x}x,k}, \quad 1 \leqslant k \leqslant N-1, \\ &y^n_0 = y^n_N = 0, \quad y^0_k = u_0(x_k), \end{aligned} \tag{28}$$

donde el parámetro $\sigma = \sigma_0 + i\sigma_1$ puede tomar valores en el plano complejo. El esquema (28) tiene un error de aproximación $O(\tau + h^2)$ para cualquier σ; cuando $\sigma = 0{,}5$, él es igual a $O(\tau^2 + h^2)$ y cuando $\sigma = 1/2 - h^2 i/(12\tau)$ el error de la aproximación es igual a $O(\tau^2 + h^4)$. La condición de estabilidad por los datos iniciales tiene la forma

$$\sigma_0 = \mathrm{Re}\,\sigma \geqslant 0{,}5. \tag{29}$$

El esquema (28) se reduce de modo usual al sistema (1″) y las condiciones del lema 1 adquieren la siguiente forma: $|2/h^2 + i(\sigma\tau)| \geqslant 2/h^2$. Resolviendo esta desigualdad, obtenemos, que el método de factorización para encontrar la solución del esquema (28) en la capa superior, al satisfacer la condición $\sigma_1 = Im\sigma \geqslant -h^2/(4\tau)$, es correcto.

De esta manera, la condición de aplicabilidad del método de factorización para el ejemplo examinado no coincide con la condición de estabilidad del mismo esquema de diferencias por los datos iniciales.

§ 2. Variantes del método de factorización

1. Variante por flujos del método de factorización. Examinemos la variante del método de factorización empleada para resolver los problemas de diferencias con coeficientes fuertemente variables. Ejemplos de tales problemas son los problemas de la hidrodinámica con termoconductividad y la hidrodinámica magnética, donde los coeficientes de termoconductividad y conductividad eléctrica dependen de los parámetros termodinámicos del medio. En el caso de problemas térmicos pueden haber partes adiabáticas donde hay ausencia de conductividad térmica y también partes isotérmicas donde la termoconductividad es infinitamente grande. En los problemas magnéticos pueden existir partes conductoras ideales y no electroconductoras, respectivamente.

Con frecuencia en tales problemas, además de la misma solución, se exige hallar también el flujo de calor (problema térmico). Al resolver las ecuaciones en diferencias de segundo orden, a las cuales se reducen los esquemas de diferencias para estos problemas, por las fórmulas de la factorización común, a menudo ocurre una pérdida significativa de exactitud. La ulterior utilización de la diferenciación numérica para calcular el flujo conduce a un resultado no satisfactorio. Librarse de esta insuficiencia se hace posible mediante el paso a la así llamada *variante por flujos del método de factorización*. Las fórmulas para esta variante de la factorización se pueden obtener como resultado de una transformación de las fórmulas de la factorización común.

Así pues, examinemos el problema de contorno de diferencias

$$-a_i y_{i-1} + c_i y_i - a_{i+1} y_{i+1} = f_i, \qquad 1 \leqslant i \leqslant N-1,$$
$$y_0 - \varkappa_1 y_1 = \mu_1, \qquad y_N - \varkappa_2 y_{N-1} = \mu_2, \tag{1}$$

donde

$$c_i = a_i + a_{i+1} + d_i, \qquad 0 \leqslant a_i < \infty, \tag{2}$$
$$d_i > 0,\ i = 1, 2, \ldots, N-1,\ |\varkappa_1| \leqslant 1,$$
$$|\varkappa_2| \leqslant 1. \tag{3}$$

Las fórmulas de factorización derecha (véase (6)—(8) del § 1) para el problema (1) tomando en consideración (2)

adquieren la forma

$$y_i = \bar{\alpha}_{i+1} y_{i+1} + \bar{\beta}_{i+1}, \quad i = N-1,\ N-2,\ \ldots,\ 0,$$

$$y_N = \frac{\mu_2 + \varkappa_2 \bar{\beta}_N}{1 - \varkappa_2 \bar{\alpha}_N}, \tag{4}$$

$$\bar{\alpha}_{i+1} = \frac{a_{i+1}}{a_{i+1} + d_i + a_i (1 - \bar{\alpha}_i)}, \quad i = 1,\ 2,\ \ldots,$$

$$N-1, \quad \bar{\alpha}_1 = \varkappa_1, \tag{5}$$

$$\bar{\beta}_{i+1} = (f_i + a_i \bar{\beta}_i) \frac{\alpha_{i+1}}{a_{i+1}}, \quad i = 1,\ 2,\ \ldots,$$

$$N-1, \quad \bar{\beta}_1 = \mu_1. \tag{6}$$

Introduzcamos una nueva función reticular desconocida (flujo) por la fórmula

$$w_i = -a_i (y_i - y_{i-1}), \qquad i = 1,\ 2,\ \ldots,\ N, \tag{7}$$

y reescribamos (1) en la forma

$$w_{i+1} - w_i + d_i y_i = f_i, \quad 1 \leqslant i \leqslant N-1,$$
$$y_0 - \varkappa_1 y_1 = \mu_1, \quad i = 0,$$
$$-\varkappa_2 w_N + a_N (1 - \varkappa_2) y_N = a_N \mu_2, \qquad i = N. \tag{8}$$

De (7) hallamos

$$y_i = y_{i-1} + \frac{1}{a_{i+1}} w_{i+1}, \quad i = 0,\ 1,\ \ldots,\ N-1,$$

y sustituyamos esta expresión en (4). Como resultado encontraremos la relación que conecta a y_{i+1} y w_{i+1}:

$$w_{i+1} + a_{i+1} (1 - \bar{\alpha}_{i+1}) y_{i+1} = a_{i+1} \bar{\beta}_{i+1},$$
$$i = 0,\ 1,\ \ldots,\ N-1.$$

Introduciendo las notaciones

$$\alpha_i = a_i (1 - \bar{\alpha}_i), \qquad \beta_i = a_i \bar{\beta}_i, \qquad i = 1,\ 2,\ \ldots,\ N,$$

reescribamos esta relación en la forma

$$w_i + \alpha_i y_i = \beta_i, \qquad i = 1,\ 2,\ \ldots,\ N. \tag{9}$$

Observemos que las ecuaciones (8), (9) forman un sistema algebraico que contiene $2N+1$ ecuaciones respecto a $2N+1$ incógnitas $y_0, y_1, \ldots, y_N$ y $w_1, w_2, \ldots, w_N$. La estructura de este sistema es tal que se divide en dos sistemas independientes para las coordenadas $y_0, y_1, \ldots$

$\ldots, y_N$ y $w_1, w_2, \ldots, w_N$. Construyamos estos sistemas.

Expresemos de (9) y_i: $y_i = (\beta_i - w_i)/\alpha_i$, $i = 1, 2, \ldots, N$ y sustituyámosla en las ecuaciones del sistema (8) para $i = 1, 2, \ldots, N$. Debido a esto obtenemos las ecuaciones

$$w_i = \frac{\alpha_i}{\alpha_i + d_i} w_{i+1} + \frac{d_i\beta_i - \alpha_i f_i}{\alpha_i + d_i},$$
$$i = N-1,\ N-2,\ \ldots,\ 1, \tag{10}$$
$$w_N = \frac{a_N\left[(1-\varkappa_2)\beta_N - \alpha_N\mu_2\right]}{(1-\varkappa_2)\,a_N + \alpha_N\varkappa_2},$$

resolviendo las cuales sucesivamente hallaremos todos los w_i.

Obtengamos ahora ecuaciones para y_i. Para esto expresemos w_i de (9): $w_i = -\alpha_i y_i + \beta_i$, $i = 1, 2, \ldots, N$ y sustituyámoslo en (8) para $i = 1, 2, \ldots, N$. Como resultado obtendremos las ecuaciones

$$y_i = \frac{\alpha_{i+1}}{\alpha_i + d_i} y_{i+1} + \frac{f_i - \beta_{i+1} + \beta_i}{\alpha_i + d_i},$$
$$i = N-1,\ N-2,\ \ldots,\ 1,$$
$$y_0 = \varkappa_1 y_1 + \mu_1, \tag{11}$$
$$y_N = \frac{\varkappa_2\beta_N + a_N\mu_2}{(1-\varkappa_2)\,a_N + \alpha_N\varkappa_2}$$

para el cálculo consecutivo de y_i.

Escribamos las fórmulas recurrentes para determinar α_i y β_i. Utilizando (5) y (6), hallamos

$$\alpha_{i+1} = a_{i+1}(1 - \bar{\alpha}_{i+1}) = \frac{a_{i+1}\left[a_i(1-\bar{\alpha}_i) + d_i\right]}{a_{i+1} + d_i + a_i(1-\bar{\alpha}_i)} =$$
$$= \frac{a_{i+1}(\alpha_i + d_i)}{a_{i+1} + \alpha_i + d_i}, \tag{12}$$
$$i = 1,\ 2,\ \ldots,\ N-1,\quad \alpha_1 = a_1(1-\varkappa_1),$$

$$\beta_{i+1} = a_{i+1}\bar{\beta}_{i+1} = \frac{a_{i+1}(f_i + \beta_i)}{a_{i+1} + \alpha_i + d_i},\quad i = 1,\ 2,\ \ldots,\ N-1,$$
$$\beta_1 = a_1\mu_1. \tag{13}$$

De las condiciones (2) y (3) y de las fórmulas (12) se deduce, que $\alpha_i \geqslant 0$. Entonces el coeficiente $\alpha_i/(\alpha_i + d_i)$ en la fórmula (10) no supera la unidad, lo cual garantiza la estabilidad del algoritmo para calcular w_i. Así sucesivamente, puesto que de las condiciones $\alpha_i \geqslant 0$ y $d_i > 0$ se desprende que $a_{i+1} < a_{i+1} + \alpha_i + d_i$, entonces en virtud de (12)

es válida la desigualdad $\alpha_{i+1} < \alpha_i + d_i$. Por eso el coeficiente $\alpha_{i+1}/(\alpha_i + d_i)$ en la fórmula (11) es siempre menor que la unidad, lo cual asegura la estabilidad al calcular y_i. Notemos que el denominador en las expresiones para w_N e y_N es siempre mayor que cero.

Así pues, el algoritmo del método de factorización por flujos se describe mediante las fórmulas (10)—(13). Observemos que es racional utilizar las fórmulas recurrentes indicadas para α_i y β_i al igual que las expresiones para y_N y w_N, si $a_{i+1} < 1$. Si $a_{i+1} \geqslant 1$, entonces se recomienda utilizar las siguientes fórmulas, obtenidas de (10)—(13) al dividir el numerador y el denominador de las fracciones por a_{i+1}:

$$\alpha_{i+1} = \frac{\alpha_i + d_i}{1 + (\alpha_i + d_i)/a_{i+1}}, \quad \beta_{i+1} = \frac{f_i + \beta_i}{1 + (\alpha_i + d_i)/a_{i+1}},$$

$$y_N = \frac{\varkappa_2 \beta_N / a_N + \mu_2}{1 - \varkappa_2 + \varkappa_2 \alpha_N / a_N}, \quad w_N = \frac{(1 - \varkappa_2)\beta_N - \alpha_N \mu_2}{1 - \varkappa_2 + \varkappa_2 \alpha_N / a_N}.$$

Calculemos el número de operaciones aritméticas que es necesario efectuar para la realización de (10)—(13). En una organización sensata de los cálculos, cuando las expresiones comunes para varias fórmulas se calculan una sola vez, y los factores comunes a varios sumandos se sacan fuera del paréntesis, el número de operaciones para (10)—(13) es $Q = 21N + 1$. Esto es aproximadamente dos veces el número de operaciones que se necesitaría gastar, para hallar la solución y_i del problema (1) por las fórmulas de factorización común, y después hallar el flujo w_i por la fórmula (7).

2. **Método de factorización cíclica.** Examinemos ahora el siguiente sistema:

$$-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = f_i,$$
$$i = 0, \pm 1, \pm 2, \ldots, \tag{14}$$

cuyos coeficientes y segundos miembros son periódicos con período N:

$$a_i = a_{i+N}, \; b_i = b_{i+N}, \; c_i = c_{i+N}, \; f_i = f_{i+N}. \tag{15}$$

A los sistemas del tipo (14), (15) llegamos, por ejemplo, al examinar esquemas de diferencias tripuntuales destinados para obtener soluciones periódicas de ecuaciones diferenciales ordinarias de segundo orden y también para la solución aproximada de ecuaciones con derivadas parciales en coordenadas cilíndricas y esféricas.

Al satisfacer las condiciones (15) la solución del sistema (14), si ella existe, también será periódica con período N,

es decir

$$y_i = y_{i+N}. \tag{16}$$

Por eso es suficiente hallar la solución y_i, por ejemplo, para $i = 0, 1, \ldots, N-1$. En este caso el problema (14)—(16) se puede escribir así:

$$-a_0 y_{N-1} + c_0 y_0 - b_0 y_1 = f_0, \qquad i = 0, \tag{17}$$

$$-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = f_i, \quad 1 \leqslant i \leqslant N-1,$$

$$y_N = y_0 \tag{18}$$

Hemos añadido la condición (18) al sistema (17), para no excluir y_N de la ecuación del sistema para $i = N-1$, sustituyéndola por y_0. Esto permite conservar una forma única para la ecuación (17) con $i = 1, 2, \ldots, N-1$.

Si introducimos los vectores de las incógnitas $Y = (y_0, y_1, \ldots, y_{N-1})$ y del segundo miembro $F = (f_0, f_1, \ldots, f_{N-1})$, entonces (17) y (18) se pueden escribir en la forma vectorial $\mathcal{A}Y = F$, donde

$$\mathcal{A} = \left\| \begin{matrix} c_0 & -b_0 & 0 & 0 & \ldots & 0 & 0 & -a_0 \\ -a_1 & c_1 & -b_1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -a_2 & c_2 & -b_2 & \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \ldots & c_{N-3} & -b_{N-3} & 0 \\ 0 & 0 & 0 & 0 & \ldots & -a_{N-2} & c_{N-2} & -b_{N-2} \\ -b_{N-1} & 0 & 0 & 0 & \ldots & 0 & -a_{N-1} & c_{N-1} \end{matrix} \right\|$$

es la matriz del sistema (17), (18). La presencia de coeficientes a_0 y b_{N-1} distintos de cero en (17), no permite resolver este sistema por el método de factorización descrito en el § 1. Para encontrar la solución del sistema (17), (18) construiremos una variante del método de factorización que se llama *método de factorización cíclica*.

Buscaremos la solución del problema (17), (18) en forma de una combinación lineal de las funciones reticulares u_i y v_i

$$y_i = u_i + y_0 v_i, \qquad 0 \leqslant i \leqslant N, \tag{19}$$

donde u_i es la solución del problema de contorno no homogéneo tripuntual

$$-a_i u_{i-1} + c_i u_i - b_i u_{i+1} = f_i, \quad 1 \leqslant i \leqslant N-1, \tag{20}$$
$$u_0 = 0, \quad u_N = 0$$

con condiciones de contorno homogéneas, y v_i es la solución del problema de contorno homogéneo tripuntual

$$-a_i v_{i-1} + c_i v_i - b_i v_{i+1} = 0, \quad 1 \leqslant i \leqslant N-1, \tag{21}$$
$$v_0 = 1, \qquad v_N = 1$$

con condiciones de contorno no homogéneas.

Hallemos bajo qué condición y_i de (19) es la solución buscada. Multiplicando (21) por y_0, sumando con (20) y teniendo en cuenta (19), obtendremos, que se cumplirán las ecuaciones del sistema (17) para $i = 1, 2, \ldots, N-1$. De las condiciones de contorno para u_i y v_i se deduce que se cumple la relación (18). De esta forma, si y_i, definida por la fórmula (19), satisface la ecuación del sistema (17) para $i = 0$ que queda sin utilizar, entonces el problema estará resuelto. Sustituyendo (19) en esta ecuación, obtendremos

$$-a_0 u_{N-1} - a_0 y_0 v_{N-1} + c_0 y_0 - b_0 u_1 -$$
$$-b_0 y_0 v_1 = f_0. \tag{22}$$

Así pues, si elegimos y_0 según la fórmula

$$y_0 = \frac{f_0 + a_0 u_{N-1} + b_0 u_1}{c_0 - a_0 v_{N-1} - b_0 v_1}, \tag{23}$$

entonces se cumplirá la igualdad (22) y, por consiguiente, la solución del problema (17), (18) se puede hallar por la fórmula (19).

Detengámonos ahora en la solución de los sistemas (20) y (21). Ellos son casos particulares de sistemas tripuntuales de ecuaciones para los cuales en el § 1 fue construido el método de factorización. Para (20) y (21) las fórmulas de factorización toman la siguiente forma:

$$u_i = \alpha_{i+1} u_{i+1} + \beta_{i+1}, \quad i = N-1, N-2, \ldots$$
$$\ldots, 1, \quad u_N = 0, \tag{24}$$
$$v_i = \alpha_{i+1} v_{i+1} + \gamma_{i+1}, \quad i = N-1, N-2, \ldots$$
$$\ldots, 1, \quad v_N = 1,$$

donde los coeficientes de factorización α_i, β_i y γ_i se encuentran por las fórmulas siguientes:

$$\alpha_{i+1} = \frac{b_i}{c_i - a_i \alpha_i}, \quad i = 1, 2, \ldots, N, \quad \alpha_1 = 0, \tag{25}$$

$$\beta_{i+1} = \frac{f_i + a_i \beta_i}{c_i - a_i \alpha_i}, \quad i = 1, 2, \ldots, N, \quad \beta_1 = 0, \tag{26}$$

$$\gamma_{i+1} = \frac{a_i \gamma_i}{c_i - a_i \alpha_i}, \quad i = 1, 2, \ldots, N, \quad \gamma_1 = 1. \tag{27}$$

Transformemos (23). De (24) obtenemos $u_{N-1} = \alpha_N u_N + \beta_N = \beta_N$, $v_{N-1} = \gamma_N + \alpha_N$. Sustituyamos estas expresiones en (23) y tengamos en cuenta las condiciones (15), (25)—(27):

$$y_0 = \frac{f_N + a_N\beta_N + \beta_N u_1}{c_N - a_N\alpha_N - \alpha_N\gamma_N - b_N v_1} = \frac{\beta_{N+1} + \alpha_{N+1}u_1}{1 - \gamma_{N+1} - \alpha_{N+1}v_1}.$$

Hemos construido el algoritmo para resolver el problema (17), (18) que lleva el nombre de método de factorización cíclica:

$$\alpha_2 = b_1/c_1, \quad \beta_2 = f_1/c_1, \quad \gamma_2 = a_1/c_1,$$

$$\alpha_{i+1} = \frac{b_i}{c_i - a_i\alpha_i}, \quad \beta_{i+1} = \frac{f_i + a_i\beta_i}{c_i - a_i\alpha_i}, \quad \gamma_{i+1} = \frac{a_i\gamma_i}{c_i - a_i\alpha_i}$$

$$i = 2, 3, \ldots, N;$$

$$u_{N-1} = \beta_N, \quad v_{N-1} = \alpha_N + \gamma_N, \tag{28}$$

$$u_i = \alpha_{i+1}u_{i+1} + \beta_{i+1}, \quad v_i = \alpha_{i+1}v_{i+1} + \gamma_{i+1},$$

$$i = N-2, N-3, \ldots, 1;$$

$$y_0 = \frac{\beta_{N+1} + \alpha_{N+1}u_1}{1 - \gamma_{N+1} - \alpha_{N+1}u_1}, \quad y_i = u_i + y_0 v_i, \quad i = 1, 2, \ldots, N-1.$$

Un cálculo elemental muestra, que para su realización se exige 6 $(N - 1)$ multiplicaciones, $5N - 3$ sumas y restos y $3N + 1$ divisiones. Si no hacemos diferencia entre las operaciones aritméticas, entonces su número general es $Q = 14N - 8$.

Investiguemos la pregunta sobre la aplicabilidad y estabilidad del algoritmo (28). Tiene lugar el siguiente lema:

LEMA 2. *Supongamos que los coeficientes del sistema* (14), (15) *satisfacen las condiciones*

$$|a_i| > 0, \quad |b_i| > 0, \quad |c_i| \geqslant |a_i| + |b_i|,$$
$$i = 1, 2, \ldots, N, \tag{29}$$

y que existe $1 \leqslant i_0 \leqslant N$ *tal que* $|c_{i_0}| > |a_{i_0}| + |b_{i_0}|$. *Entonces*

$$c_i - a_i\alpha_i \neq 0, \quad |\alpha_i| \leqslant 1, \quad |\alpha_i| + |\gamma_i| \leqslant 1,$$
$$i = 2, 3, \ldots, N,$$

$$1 - \gamma_{N+1} - \alpha_{N+1}v_1 \neq 0.$$

En efecto, ya que α_i, β_i y γ_i son los coeficientes de factorización del método de factorización derecha, aplicado a la solución de los problemas (20) y (21), y en virtud de (29) están cumplidas las condiciones del lema 1, entonces

del lema 1 se desprende la justeza de las desigualdades

$$c_i - a_i\alpha_i \neq 0, \quad |\alpha_i| \leqslant 1, \quad i = 2, 3, \ldots, N,$$
$$|c_i - a_i\alpha_i| \geqslant |c_i| - |a_i|\,|\alpha_i| \geqslant |b_i| > 0. \tag{30}$$

A continuación a base de las condiciones del lema 2, $|a_1| + |b_1| \leqslant |c_1|$ y, por lo tanto, $|\alpha_2| + |\gamma_2| \leqslant 1$. De aquí por el método de inducción obtenemos las desigualdades

$$|\alpha_i| + |\gamma_i| \leqslant 1, \quad i = 2, 3, \ldots, N, \tag{31}$$

como

$$|\alpha_{i+1}| + |\gamma_{i+1}| = \frac{|b_i| + |a_i|\,|\gamma_i|}{|c_i - a_i\alpha_i|} \leqslant$$
$$\leqslant \frac{|a_i| + |b_i| - |a_i|(1 - |\gamma_i|)}{|c_i| - |a_i|\,|\alpha_i|} \leqslant \frac{|a_i| + |b_i| - |a_i|\,|\alpha_i|}{|c_i| - |a_i|\,|\alpha_i|} \leqslant 1,$$

y se tiene (30).
Notemos, que $|c_i| > |a_i| + |b_i|$ para $i = i_0$ y, por tanto, $|\alpha_{i_0+1}| + |\gamma_{i_0+1}| < 1$. De aquí se deduce que para $i \geqslant i_0 + 1$ tiene lugar la desigualdad estricta $|\alpha_i| + |\gamma_i| < 1$. Como $1 \leqslant i_0 \leqslant N$, entonces $|\alpha_{N+1}| + |\gamma_{N+1}| < 1$.

Nos queda por mostrar que $1 - \gamma_{N+1} - \alpha_{N+1}v_1 \neq 0$. A base de (28) y (31) obtenemos

$$|v_{N-1}| \leqslant |\alpha_N| + |\gamma_N| \leqslant 1,$$

y más adelante por el método de inducción demostramos las desigualdades $|v_i| \leqslant 1$, $1 \leqslant i \leqslant N - 1$, ya que en virtud de (31)

$$|v_i| \leqslant |\alpha_{i+1}|\,|v_{i+1}| + |\gamma_{i+1}| \leqslant |\alpha_{i+1}| + |\gamma_{i+1}| \leqslant 1.$$

En particular, $|v_1| \leqslant 1$. De aquí, teniendo en cuenta la desigualdad demostrada $|\alpha_{N+1}| + |\gamma_{N+1}| < 1$, sacamos la conclusión de que

$$|1 - \gamma_{N+1} - \alpha_{N+1}v_1| \geqslant 1 - |\gamma_{N+1}| - |\alpha_{N+1}|\,|v_1| \geqslant$$
$$\geqslant 1 - |\alpha_{N+1}| - |\gamma_{N+1}| > 0.$$

El lema 2 está completamente demostrado.

Como conclusión notemos, que el coeficiente de factorización β_i, depende de la parte derecha f_i, y por consiguiente, de u_i e y_i. Los coeficientes de factorización α_i y γ_i al igual

que v_i no dependen de f_i y se calculan y se guardan en la memoria al resolver tan sólo el primer problema de la serie. Esto permite resolver el segundo y cada problema siguiente de la serie mediante $Q = 9N - 4$ operaciones.

3. Método de factorización para sistemas complejos. Continuemos la construcción de variantes del método de factorización para la solución de sistemas de ecuaciones de diferencias con matrices distintas de las tridiagonales. En el punto 2 se aplicó el método de factorización cíclica para resolver sistemas cuyas matrices contenían solamente dos elementos no nulos fuera de las diagonales principales. Examinemos ahora un caso más general.

Supongamos que se exige resolver el siguiente sistema de ecuaciones:

$$\begin{aligned} &c_0 y_0 - \sum_{j=1}^{N-1} d_j y_j - \psi_0 y_N = f_0, \quad i = 0,\\ &-\varphi_i y_0 - a_i y_{i-1} + c_i y_i - b_i y_{i+1} - \psi_i y_N = f_i, \quad 1 \leqslant i \leqslant N-1, \qquad (32)\\ &-\varphi_N y_0 \sum_{j=1}^{N-1} g_j y_j + c_N y_N = f_N, \quad i = N. \end{aligned}$$

El sistema del tipo (32) aparece al aproximar las ecuaciones diferenciales ordinarias de segundo orden en el caso de condiciones de contorno relacionadas, al encontrar las soluciones que satisfacen condiciones complementarias de tipo integral, y en una serie de otros problemas. En particular se pueden escribir en esta forma todos los sistemas de ecuaciones de diferencias examinadas más arriba. Por ejemplo, si en (32) ponemos

$$\begin{aligned} &d_1 = b_0, \quad d_{N-1} = a_0, \quad d_i = 0, \quad 2 \leqslant i \leqslant N-2,\\ &\varphi_i = \psi_i = g_i = 0, \quad 1 \leqslant i \leqslant N-1,\\ &\psi_0 = 0, \quad \varphi_N = c_N = 1 \quad f_N = 0, \end{aligned}$$

entonces obtendremos el problema (17), (18).

Si introducimos los vectores $Y = (y_0, y_1, \ldots, y_N)$ y $F = (f_0, \ldots, f_N)$, entonces (32) se puede escribir en la forma vectorial $\mathcal{A}Y = F$, donde

$$\mathcal{A} = \left\| \begin{array}{c:cccccccc:c} c_0 & -d_1 & -d_2 & -d_3 & \ldots & -d_{N-3} & -d_{N-2} & -d_{N-1} & -\psi_0 \\ \hdashline -\varphi_1 - a_1 & c_1 & -b_1 & 0 & \ldots & 0 & 0 & 0 & -\psi_1 \\ -\varphi_2 & -a_2 & c_2 & -b_2 & \ldots & 0 & 0 & 0 & -\psi_2 \\ -\varphi_3 & 0 & -a_3 & c_3 & \ldots & 0 & 0 & 0 & -\psi_3 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ -\varphi_{N-3} & 0 & 0 & 0 & \ldots & c_{N-3} & -b_{N-3} & 0 & -\psi_{N-3} \\ -\varphi_{N-2} & 0 & 0 & 0 & \ldots & -a_{N-2} & c_{N-2} & -b_{N-2} & -\psi_{N-2} \\ -\varphi_{N-1} & 0 & 0 & 0 & \ldots & 0 & -a_{N-1} & c_{N-1} & -b_{N-1} - \psi_{N-1} \\ \hdashline -\varphi_N & -g_1 & -g_2 & -g_3 & \ldots & -g_{N-3} & -g_{N-2} & -g_{N-1} & c_N \end{array} \right\|$$

Se ve que la matriz $\mathcal{A}$ se obtiene bordeando una matriz tridiagonal con ayuda de columnas y filas desde todos los cuatro lados. Observemos

que para otro ordenamiento de las incógnitas $Y^* = (y_1, y_2, \ldots, y_N, y_0)$ el sistema (32) se escribe en la forma $\mathcal{A}^* Y^* = F^*$, donde la matriz $\mathcal{A}^*$ se obtiene al bordear la misma matriz tridiagonal, pero solamente con ayuda de dos columnas a la derecha y de dos filas de abajo.

Pasemos a la construcción de un método de solución del problema (32). Buscaremos la solución del problema (32) en forma de una combinación lineal de tres funciones reticulares u_i, v_i y w_i:

$$y_i = u_i + y_0 v_i + y_N w_i, \qquad 0 \leqslant i \leqslant N, \tag{33}$$

donde u_i, v_i y w_i son las soluciones de los siguientes problemas de contorno tripuntuales:

$$\left.\begin{array}{l} -a_i u_{i-1} + c_i u_i - b_i u_{i+1} = f_i, \quad 1 \leqslant i \leqslant N-1, \\ u_0 = 0, \quad u_N = 0; \end{array}\right\} \tag{34}$$

$$\left.\begin{array}{l} -a_i v_{i-1} + c_i v_i - b_i v_{i+1} = \varphi_i, \quad 1 \leqslant i \leqslant N-1, \\ v_0 = 1, \quad v_N = 0; \end{array}\right\} \tag{35}$$

$$\left.\begin{array}{l} -a_i w_{i-1} + c_i w_i - b_i w_{i+1} = \psi_i, \quad 1 \leqslant i \leqslant N-1, \\ w_0 = 0, \quad w_N = 1. \end{array}\right\} \tag{36}$$

De (33)-(36) se ve, que para $1 \leqslant i \leqslant N-1$ se cumplen las ecuaciones del sistema (32). Las condiciones de contorno para u_i, v_i y w_i garantizan la reducción de (33) a una identidad para $i = 0$ o $i = N$. De esta manera, si se resolverán los problemas (34)—(36) y estarán halladas y_0 y y_N, la fórmula (33) definirá la solución del problema inicial (32). Hallemos primeramente y_0 e y_N.

Hallaremos los valores para y_0 e y_N, empleando las ecuaciones del sistema (32) si $i = 0$ e $i = N$. Sustituyendo y_i de (33) en estas ecuaciones, obtenemos un sistema de dos ecuaciones para y_0 e y_N:

$$\Big(c_0 - \sum_{j=1}^{N-1} d_j v_j\Big) y_0 - \Big(\psi_0 + \sum_{j=1}^{N-1} d_j w_j\Big) y_N = f_0 + \sum_{j=1}^{N-1} d_j u_j,$$

$$-\Big(\varphi_N + \sum_{j=1}^{N-1} g_j v_j\Big) y_0 + \Big(c_N - \sum_{j=1}^{N-1} g_j w_j\Big) y_N = f_N + \sum_{j=1}^{N-1} g_j u_j.$$

Si el determinante de este sistema

$$\Delta = \Big(c_0 - \sum_{j=1}^{N-1} d_j v_j\Big)\Big(c_N - \sum_{j=1}^{N-1} g_j w_j\Big) - \\ -\Big(\psi_0 + \sum_{j=1}^{N-1} d_j w_j\Big)\Big(\varphi_N + \sum_{j=1}^{N-1} g_j v_j\Big) \tag{37}$$

des distinto del cero, entonces el sistema tiene la única solución

$$y_0 = \frac{1}{\Delta}\Big[\Big(c_N - \sum_{j=1}^{N-1} g_j w_j\Big)\Big(f_0 + \sum_{j=1}^{N-1} d_j u_j\Big) + \\ + \Big(\psi_0 + \sum_{j=1}^{N-1} d_j w_j\Big)\Big(f_N + \sum_{j=1}^{N-1} g_j u_j\Big)\Big], \tag{38}$$

$$y_N=\frac{1}{\Delta}\left[\left(\varphi_N+\sum_{j=1}^{N-1} g_j v_j\right)\left(f_0+\sum_{j=1}^{N-1} d_j u_j\right)+\right.$$

$$\left.+\left(c_0-\sum_{j=1}^{N-1} d_j v_j\right)\left(f_N+\sum_{j=1}^{N-1} g_j u_j\right)\right]. \qquad (39)$$

Examinemos ahora el método de solución de los problemas auxiliares (34)-(36). Ya que aquí nos tropezamos con problemas de contorno ordinarios para ecuaciones tripuntuales, entonces se puede utilizar el método de factorización descrito en el § 1. Para (34)-(36) las fórmulas del algoritmo de factorización derecha adquieren la siguiente forma:

$$\begin{aligned} u_i &= \alpha_{i+1}u_{i+1}+\beta_{i+1}, \quad i=N-1,\ \ldots,\ 0,\quad u_N=0,\\ v_i &= \alpha_{i+1}v_{i+1}+\gamma_{i+1}, \quad i=N-1,\ \ldots,\ 0,\quad v_N=0, \qquad (40)\\ w_i &= \alpha_{i+1}w_{i+1}+\delta_{i+1}, \quad i=N-1,\ \ldots,\ 0,\quad w_N=1, \end{aligned}$$

donde los coeficientes de factorización α_i, $\beta_i \gamma_i$ y δ_i se determinan por las fórmulas

$$\begin{aligned} &\alpha_{i+1}=\frac{b_i}{c_i-a_i\alpha_i}, \quad \beta_{i+1}=\frac{f_i+a_i\beta_i}{c_i-a_i\alpha_i},\\ &i=1,\ 2,\ \ldots,\ N-1, \quad \alpha_1=0,\ \beta_1=0, \qquad (41)\\ &\gamma_{i+1}=\frac{\varphi_i+a_i\gamma_i}{c_i-a_i\alpha_i}, \quad \delta_{i+1}=\frac{\psi_i+a_i\delta_i}{c_i-a_i\alpha_i},\\ &i=1,\ 2,\ \ldots,\ N-1, \quad \gamma_1=1, \quad \delta_1=0. \end{aligned}$$

Por lo tanto, para el problema (32) el método de factorización se describe por las fórmulas (33), (37)-(41).

Examinemos ahora la pregunta sobre la estabilidad y corrección del algoritmo propuesto. En virtud del lema 1 las condiciones

$$\begin{aligned} &|a_i|>0, \quad |b_i|>0, \quad |c_i|\geqslant|a_i|+|b_i|,\\ &1\leqslant i\leqslant N-1 \qquad (42) \end{aligned}$$

son suficientes para la estabilidad y corrección del método de factorización (40)-(41) de solución de los problemas auxiliares (34)-(36). Se puede mostrar, que si el sistema inicial (32) tiene solución única, entonces el determinante Δ, definido por la fórmula (37), es distinto de cero. En este caso las fórmulas (38) y (39) para calcular y_0 y y_N serán correctas. Formulemos el resultado obtenido en forma de un lema.

LEMA 3. *Si el sistema (32) tiene solución única y son cumplidas las condiciones (42), entonces el algoritmo (33), (37)-(41) del método de factorización para el problema (32) es correcto y estable.*

Observemos que la formulación de las sencillas y a la vez no muy restrictivas condiciones suficientes de solubilidad del sistema (32) es un problema complejo. Citemos un ejemplo de las condiciones que aseguran la corrección y la estabilidad del algoritmo propuesto. Supongamos que la matriz del sistema (32) posee una predominancia

diagonal, es decir, se cumplen las condiciones

$$|c_i| \geqslant |a_i| + |b_i| + |\varphi_i| + \psi_i, \quad 1 \leqslant i \leqslant N-1, \tag{43}$$

$$|c_0| \geqslant |\psi_0| + \sum_{j=1}^{N-1} |d_j|, \quad |c_N| \geqslant |\varphi_N| + \sum_{j=1}^{N-1} |g_j|, \tag{44}$$

$$|a_i| > 0, \ |b_i| > 0, \ 1 \leqslant i \leqslant N-1, \ |c_0| > 0, \ |c_N| > 0,$$

y además, al menos en una de las desigualdades (43) o (44) se cumple la desigualdad estricta.

Indiquemos las etapas fundamentales de la demostración. Al principio se demuestra que tienen lugar las desigualdades $|\alpha_i| + |\gamma_i| + |\delta_i| \leqslant 1$, $1 \leqslant i \leqslant N$. A continuación se demuestran las desigualdades $|v_i| + |w_i| \leqslant 1$, para $1 \leqslant i \leqslant N$, además, si en (43) al menos para un i se cumple la desigualdad estricta, entonces para todos $1 \leqslant i \leqslant N$ son ciertas las desigualdades $|v_i| + |w_i| < 1$. Seguidamente tenemos

$$\left|c_0 - \sum_{j=1}^{N-1} d_j v_j\right| \geqslant |c_0| - \sum_{j=1}^{N-1} |d_j|\,|v_j| \geqslant |\psi_0| +$$

$$+ \sum_{j=1}^{N-1} (1 - |v_j|)\,|d_j| \geqslant |\psi_0| + \sum_{j=1}^{N-1} |w_j|\,|d_j| \geqslant \left|\psi_0 \sum_{j=1}^{N-1} w_j d_j\right|$$

y análogamente

$$\left|c_N - \sum_{j=1}^{N-1} g_j w_j\right| \geqslant \left|\varphi_N + \sum_{j=1}^{N-1} g_j v_j\right|,$$

pero al menos en una de estas desigualdades se alcanza la desigualdad estricta. De aquí se deduce, que el determinante Δ definido en (37) es distinto de cero. La estabilidad y corrección del método de factorización para la solución de los problemas auxiliares (34)-(36) se desprenden de (43).

En calidad de ejemplo de un problema reducible a (32), examinemos el esquema con pesos

$$\begin{aligned} & y_{t,i} = \sigma y^{n+1}_{\bar{x}x,i} + (1-\sigma)\, y^{n}_{\bar{x}x,i}, \quad 1 \leqslant i \leqslant N-1, \\ & y^n_0 - y^n_k = \mu_1(t_n), \quad y^n_N - y^n_k = (\mu^2(t_n), \\ & y^0_i = u_0(x_i), \quad n = 0, 1, \ldots, \quad 1 \leqslant k \leqslant N-1, \end{aligned} \tag{45}$$

que aproxima la ecuación de conducción del calor con condiciones de contorno relacionadas no locales

$$\begin{aligned} & \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < l, \ t > 0, \\ & u(0, t) - u(\nu(t), t) = \mu_1(t), \\ & u(l, t) - u(\nu(t), t) = (\mu_2(t), \quad u(x, 0) = u_0(x), \end{aligned}$$

donde la función $x = \nu(t)$ toma valores de 0 hasta l. Notemos que en el esquema (45) la curva $x = \nu(t)$ es aproximada por la quebrada $x_k = \nu(t_n)$, de manera tal que los puntos (x_k, t_n) son los puntos de empalme de la red.

El esquema de diferencias (45) se escribe en la forma del sistema (32) respecto a $y_i = y_i^{n+1}$ para los siguientes valores de los coeficientes y del segundo miembro ($\sigma \neq 0$):

$$c_0 = 1,\ d_k = 1,\ f_0 = \mu_1(t_{n+1}),\ \psi_0 = 0,\ d_j = 0,\ j \neq k,$$
$$c_N = 1,\ q_k = 1,\ f_N = \mu_2(t_{n+1}),\ \varphi_N = 0,\ g_j = 0,\ j \neq k,$$
$$\varphi_j = \psi_j = 0,\ a_i = b_i = 1/h^2,\ c_i = a_i + b_i + 1/(\sigma\tau),$$
$$f_i = \frac{1}{\sigma\tau} y_i^n + \left(\frac{1}{\sigma} - 1\right) y_{\bar{x}x,\,i}^n,\quad i = 1, 2, \ldots, N-1.$$

De aquí obtenemos que la exigencia $|2/h^2 + 1/(\sigma\tau)| > 2/h^2$ asegura el cumplimiento de las condiciones (43), (44). Por consiguiente, para encontrar la solución del esquema (45) sobre la capa superior si $\sigma > -h^2/(4\tau)$, se puede utilizar la variante aquí descrita del método de factorización, la cual será estable y correcta.

4. Método de factorización no monótona. Regresemos de nuevo al método de factorización para la solución de acuaciones tripuntuales:

$$\begin{aligned} &c_0 y_0 - b_0 y_1 = f_0, \quad i = 0, \\ &-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = f_i, \qquad i = 1, 2, \ldots, N-1, \\ &-a_N y_{N-1} + c_N y_N = f_N, \qquad i = N, \end{aligned} \tag{46}$$

que fue construido en el § 1. Recordemos que en el algoritmo de factorización derecha (izquierda) las incógnitas y_i se encuentran sucesivamente al moverse en la dirección de disminución (de aumento) del índice i. Al mismo tiempo y_i se expresa únicamente mediante la incógnita vecina. Tal estructura del algoritmo nos da fundamento para llamar al método construido, *método de factorización no monótona*.

El orden monótono de definición de las incógnitas y_i en el paso inverso del método está proporcionado por el orden natural de eliminación de las incógnitas de las ecuaciones durante el paso directo. De esta forma, el método de factorización monótona es el método de eliminación de Gauss sin elegir el elemento principal, aplicado al sistema especial de ecuaciones algebraicas lineales (46) con matriz tridiagonal. Es conocido, que esta variante del método de eliminación de Gauss es correcta para el caso de sistema de ecuaciones con matrices que poseen predominancia diagonal. Para el sistema (46) esta afirmación está demostrada en el lema 1.

Detengámonos en esto más detalladamente. Recordemos que en el § 1, punto 1, en el l-ésimo paso del proceso de eliminación de las incógnitas en el sistema (46) fue obtenido

el sistema «abreviado»

$$(c_l - a_l\alpha_l)\, y_l - b_l y_{l+1} = f_l + a_l\beta_l, \qquad i = l,$$
$$-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = f_i, \qquad l + 1 \leqslant i \leqslant N - 1, \quad (47)$$
$$-a_N y_{N-1} + c_N y_N = f_N$$

para las incógnitas $y_l, y_{l+1}, \ldots, y_N$. Suponiendo que $c_l - a_l\alpha_l$ es distinto de cero, nosotros transformamos la primera ecuación del sistema (47) a la forma

$$y_l = \alpha_{l+1} y_{l+1} + \beta_{l+1}, \quad \alpha_{l+1} = b_l/(c_l - a_l\alpha_l) \qquad (48)$$

y la utilizamos para excluir y_l de la ecuación (47) para $i = l + 1$. El lema 1 afirma que si la matriz $\mathcal{A}$ del sistema (46) tiene una predominancia diagonal, entonces es válida la desigualdad $|c_l - a_l\alpha_l| \geqslant |b_l|$. Por consiguiente, en la primera ecuación del sistema (47) el coeficiente de y_l es mayor en módulo que el coeficiente de y_{l+1}. Por eso no es necesario efectuar la elección del elemento principal por la fila, la transición a la forma (48) es correcta y la condición de estabilidad $|\alpha_{l+1}| \leqslant 1$ se cumple automáticamente.

Si no tiene lugar la predominancia diagonal, entonces no se puede garantizar que las cantidades $c_l - a_l\alpha_l$ sean distintas de cero así como la desigualdad $|\alpha_{l+1}| \leqslant 1$. En este caso el algoritmo de factorización monótona puede proporcionar la división por cero o una fuerte sensibilidad a los errores de redondeo, y por consiguiente, se debe modificar este algoritmo. La construcción de un algoritmo correcto del método de factorización para el sistema (46), que posea una solución única, se basa en emplear la elección de un elemento principal por filas en el método de eliminación de Gauss. En dicho algoritmo puede ser violado el orden monótono de definición de las incógnitas y_i y por lo tanto este método lo llamaremos *método de factorización no monótona*.

Pasemos a la descripción del algoritmo de la factorización no monótona. Supongamos que debido al l-ésimo paso del proceso de eliminación de Gauss con la elección del elemento principal por fila, aplicado al sistema (46), se obtiene el siguiente sistema «reducido»:

$$C y_{m_l} - b_l y_{l+1} = F, \qquad i = l, \qquad (49)$$
$$-A y_{m_l} + c_{l+1} y_{l+1} - b_{l+1} y_{l+2} = \Phi, \quad i = l+1, \quad (50)$$
$$-a_{l+2} y_{l+1} + c_{l+2} y_{l+2} - b_{l+2} y_{l+3} = f_{l+2},$$
$$i = l + 2, \qquad (51)$$

$$-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = f_i, \quad l+3 \leqslant i \leqslant N-1, \tag{52}$$

$$-a_N y_{N-1} + c_N y_N = f_N, \quad i = N, \tag{53}$$

donde $m_l \leqslant l$. (Para $l = 0$ en (49)—(53) se debe poner $C = c_0$, $A = a_1$, $F = f_0$, $\Phi = f_1$ y $m_0 = 0$).

Describamos el $l+1$-ésimo paso del proceso de eliminación. La estrategia de elección del elemento principal por fila nos conduce a dos casos:

a) Si $|C| \geqslant |b_l|$, la ecuación (49) se transforma en la forma

$$y_{m_l} - \alpha_{l+1} y_{l+1} = \beta_{l+1}, \quad \alpha_{l+1} = b_l/C, \quad \beta_{l+1} = F/C,$$

pero $|\alpha_{l+1}| \leqslant 1$ y la incógnita con índice m_l se encuentra mediante la incógnita con índice $l+1$. Así sucesivamente, con ayuda de la ecuación obtenida se elimina y_{m_l} de (50). Esto da la siguiente ecuación:

$$Cy_{m_{l+1}} - b_{l+1} y_{l+2} = F, \quad i = l+1, \tag{54}$$

donde hemos denotado $m_{l+1} = l+1$, $C = c_{l+1} - A\alpha_{l+1}$, $F = \Phi + A\beta_{l+1}$. La ecuación (51) no se transforma, ya que ella no contiene y_{m_l}, pero se vuelve a escribir en la forma

$$-Ay_{m_{l+1}} + C_{l+2} y_{l+2} - b_{l+2} y_{l+3} = \Phi, \quad i = l+2, \tag{55}$$

donde se supone $A = a_{l+2}$ y $\Phi = f_{l+2}$. Uniendo (54) y (55) con (52) y (53), obtenemos un nuevo sistema «reducido» del tipo (49)—(53), en el cual l se ha sustituido por $l+1$. Con esto se termina el $(l+1)$-ésimo paso.

b) Si $|C| < |b_l|$, entonces (49) se transforma a la forma

$$y_{l+1} - \alpha_{l+1} y_{m_l} = \beta_{l+1}, \ \alpha_{l+1} = C/b_l, \ \beta_{l+1} = -F/b_l,$$

donde de nuevo $|\alpha_{l+1}| \leqslant 1$, sin embargo esta vez la incógnita con índice $l+1$ se calcula por medio de la incógnita con índice m_l. La ecuación obtenida se utiliza para eliminar y_{l+1} de (50) y (51). Con esto la ecuación (50) se habrá transformado a la forma (54), donde $m_{l+1} = m_l$, $C = c_{l+1}\alpha_{l+1} - A$, $F = \Phi - c_{l+1}\beta_{l+1}$ y la ecuación (51) a la forma (55), donde las cantidades A y Φ se redefinen por las fórmulas $A = a_{l+2}\alpha_{l+1}$, $\Phi = f_{l+2} + a_{l+2}\beta_{l+1}$. Las ecuaciones (52),

(53) no se transforman, ya que no contienen y_{i+1}. De nuevo obtenemos un sistema del tipo (49)—(53). El se diferencia del obtenido en el primer caso por los coeficientes C y A y los miembros derechos F y Φ calculados según las otras fórmulas.

Así pues, hemos descrito un paso del proceso de eliminación con elección del elemento principal. Notemos que si el sistema inicial no es degenerado, en la ecuación (49) los coeficientes C y b_i no pueden anularse simultáneamente. Esto asegura la corrección de las fórmulas para los coeficientes de factorización α_{i+1} y β_{i+1}. Como todos los α_{i+1} calculados no exceden en módulo la unidad, entonces el proceso de cálculo de las incógnitas y_i en el paso inverso del método será estable respecto a los errores de redondeo.

Para el algoritmo propuesto el orden del cálculo de las incógnitas puede tener un carácter no monótono. Esto exige que se conserve la información acerca de cual incógnita se calcula y a través de la cual de las ya halladas en los pasos anteriores, con ayuda de los coeficientes de factorización α_{i+1} y β_{i+1}. Esta información se puede conservar en forma de dos conjuntos de índices θ y $\varkappa$ de números enteros: $\theta = \{\theta_i, 1 \leqslant i \leqslant N\}$, $\varkappa = \{\varkappa_i, 1 \leqslant i \leqslant N\}$, puesto que las incógnitas se encuentran por las fórmulas $y_m = \alpha_{i+1}y_n + \beta_{i+1}$, $m = \theta_{i+1}$, $n = \varkappa_{i+1}$, $i = N-1, N-2, \ldots, 0$. Los conjuntos θ y $\varkappa$ se construyen en el paso directo del método.

El algoritmo del método de factorización no monótona se puede definir de la siguiente forma:

1) Se dan los valores iniciales para C, A, F y Φ: $C = c_0$, $A = a_1$, $F = f_0$, $\Phi = f_1$ y formalmente se supone $\varkappa_0 = 0$.

2) Sucesivamente para $i = 0, 1, \ldots, N-1$ se efectúan, en función de la situación, las operaciones descritas en los puntos a) o b):

a) si $|C| \geqslant |b_i|$, entonces $\alpha_{i+1} = b_i/C$, $\beta_{i+1} = F/C$, $C = c_{i+1} - A\alpha_{i+1}$, $F = \Phi + A\beta_{i+1}$, $\theta_{i+1} = \varkappa_i$, $\varkappa_{i+1} = i + 1$, $A = a_{i+2}$, $\Phi = f_{i+2}$;

b) si $|C| < |b_i|$, entonces $\alpha_{i+1} = C/b_i$, $\beta_{i+1} = -F/b_i$; $C = c_{i+1}\alpha_{i+1} - A$, $F = \Phi - c_{i+1}\beta_{i+1}$, $\theta_{i+1} = i + 1$, $\varkappa_{i+1} = \varkappa_i$, $A = a_{i+2}\alpha_{i+1}$, $\Phi = f_{i+2} + a_{i+2}\beta_{i+1}$.

OBSERVACIÓN. Para $i = N-1$ no es necesario redefinir A y Φ en los puntos a) y b).

3) Se calcula inicialmente la incógnita y_n, donde $n = \varkappa_N$ por la fórmula $y_n = F/C$, y después sucesivamente

para $i = N - 1, N - 2, \ldots, 0$ se calculan las restantes incógnitas $y_m = \alpha_{i+1} y_n + \beta_{i+1}$, $m = \theta_{i+1}$, $n = \varkappa_{i+1}$.

Notemos que el algoritmo aquí propuesto pasa a ser el algoritmo ordinario de factorización derecha, si se satisfacen las condiciones del lema 1.

Un cálculo elemental del número de operaciones aritméticas para el algoritmo del método de factorización no monótona muestra, que en el peor de los casos, cuando para cualquier i los cálculos se realizan por las fórmulas del punto b), se exigen $Q = 12N$ operaciones. Esto es en 1,5 veces mayor que en el algoritmo de factorización monótona.

Examinemos un ejemplo de aplicación del método de factorización no monótona. Supongamos que se exige resolver el siguiente problema de diferencias:

$$-y_{i-1} + y_i - y_{i+1} = 0, \quad 1 \leqslant i \leqslant N - 1,$$
$$y_0 = 1, \quad y_N = 0. \tag{56}$$

El problema (56) es un caso particular del sistema (46), en el cual $f_0 = 1$, $b_0 = a_N = 0$, $c_0 = c_N = 1$, $f_N = 0$, $c_i = a_i = b_i = 1$, $f_i = 0$, $1 \leqslant i \leqslant N - 1$. Si N no es múltiplo de 3, la solución del problema (56) existe y tiene la forma (véase el punto 1 del § 4 del cap. I)

$$y_i = \operatorname{sen} \frac{(N-i)\pi}{3} \Big/ \operatorname{sen} \frac{N\pi}{3}, \quad 0 \leqslant i \leqslant N. \tag{57}$$

Tabla 1

i	0	1	2	3	4	5	6	7	8	9	10	11
α_i		0	1	0	−1	1	0	−1	1	0	−1	1
β_i		1	1	−1	−1	−1	1	1	1	−1	−1	−1
θ_i		0	1	3	2	4	6	5	7	9	8	10
$\varkappa_i$		1	2	2	4	5	5	7	8	8	10	11
y_i	1	1	0	−1	−1	0	1	1	0	−1	−1	0

Los algoritmos de factorización derecha o izquierda para (56) no son correctos, ya que al calcular los coeficientes de factorización α_3 para la factorización derecha y ξ_{N-2} para la factorización izquierda sería necesario dividir por el denominador nulo $c_2 - a_2\alpha_2$ ó $c_{N-2} - b_{N-2}\xi_{N-1}$. El algoritmo de la factorización no monótona permite obtener la solución exacta (57). Citemos para ilustrar (tabla 1) los valores de los coeficientes α_i, β_i y además θ_i y $\varkappa_i$ para $N = 11$.

§ 3. Método de factorización para ecuaciones pentapuntuales

1. Algoritmo de factorización monótona. Más arriba examinamos diferentes variantes del método de factorización que se aplican para encontrar la solución de ecuaciones de diferencias tripuntuales. Como fue señalado anteriormente, estas ecuaciones de diferencias aparecen al aproximar los problemas de contorno para ecuaciones diferenciales ordinarias de segundo orden.

Para encontrar la solución de los problemas de contorno para ecuaciones de orden más alto se pueden utilizar dos métodos. El primer método consiste en la transición a un sistema de ecuaciones diferenciales de primer orden y la construcción del correspondiente esquema de diferencias. En este caso obtendremos un problema de contorno para ecuaciones vectoriales bipuntuales. Los métodos de solución de tales problemas de diferencias serán examinadas en el § 4.

El segundo método consiste en la aproximación directa del problema diferencial. En este caso llegamos a ecuaciones de diferencias multipuntuales. Con más frecuencia se encuentran los sistemas de ecuaciones pentapuntuales del siguiente tipo:

$$c_0y_0 - d_0y_1 + e_0y_2 = f_0, \qquad i = 0, \quad (1)$$

$$-b_1y_0 + c_1y_1 - d_1y_2 + e_1y_3 = f_1, \qquad i = 1, \quad (2)$$

$$a_iy_{i-2} - b_iy_{i-1} + c_iy_i - d_iy_{i+1} + e_iy_{i+2} = f_i, \qquad 2 \leqslant i \leqslant N-2, \quad (3)$$

$$a_{N-1}y_{N-3} - b_{N-1}y_{N-2} + c_{N-1}y_{N-1} - d_{N-1}y_N = f_{N-1}, \qquad i = N-1, \quad (4)$$

$$a_Ny_{N-2} - b_Ny_{N-1} + c_Ny_N = f_N, \qquad i = N. \quad (5)$$

Este tipo de sistemas aparece al aproximar los problemas de contorno para ecuaciones diferenciales ordinarias de cuarto orden, y también en la realización de los esquemas de diferencias para ecuaciones en derivadas parciales. La matriz $\mathcal{A}$ del sistema (1)—(5) es una matriz cuadrada pentadiagonal de dimensión $(N+1) \times (N+1)$ y posee no más de $5N - 1$ elementos no nulos.

Para resolver el sistema (1)—(5) utilizaremos el método de eliminación de Gauss. Teniendo en cuenta la estructura del sistema (1)—(5), obtenemos fácilmente, que el paso inverso del método de Gauss debe realizarse según las fórmulas

$$y_i = \alpha_{i+1} y_{i+1} - \beta_{i+1} y_{i+2} + \gamma_{i+1},$$
$$0 \leqslant i \leqslant N-2, \tag{6}$$
$$y_{N-1} = \alpha_N y_N + \gamma_N, \quad i = N-1. \tag{7}$$

Para la realización de (6) y (7) es preciso prefijar y_N, y además determinar los coeficientes α_i, β_i y γ_i.

Primeramente hallemos las fórmulas para α_i, β_i y γ_i. Aplicando (6), expresemos y_{i-1} e y_{i-2} por medio de y_i e y_{i+1}. Obtendremos

$$y_{i-1} = \alpha_i y_i - \beta_i y_{i+1} + \gamma_i, \quad 1 \leqslant i \leqslant N-1, \tag{8}$$
$$y_{i-2} = (\alpha_i \alpha_{i-1} - \beta_{i-1}) y_i - \beta_i \alpha_{i-1} y_{i+1} + \\ + \alpha_{i-1}\gamma_i + \gamma_{i-1} \tag{9}$$

para $2 \leqslant i \leqslant N-1$.

Sustituyendo (8) y (9) en (3), obtenemos

$$[c_i - a_i\beta_{i-1} + \alpha_i (a_i\alpha_{i-1} - b_i)] y_i = \\ = [d_i + \beta_i (a_i\alpha_{i-1} - b_i)] y_{i+1} - e_i y_{i+2} + \\ + [f_i - a_i\gamma_{i-1} - \gamma_i (a_i\alpha_{i-1} - b_i)],$$
$$2 \leqslant i \leqslant N-2.$$

Comparando esta expresión con (6), vemos que si se pone

$$\alpha_{i+1} = \frac{1}{\Delta_i}[d_i + \beta_i (a_i\alpha_{i-1} - b_i)], \quad \beta_{i+1} = \frac{e_i}{\Delta_i}, \tag{10}$$
$$\gamma_{i+1} = \frac{1}{\Delta_i}[f_i - a_i\gamma_{i-1} - \gamma_i (a_i\alpha_{i-1} - b_i)],$$

donde se designa $\Delta_i = c_i - a_i\beta_{i-1} + \alpha_i (a_i\alpha_{i-1} - b_i)$, entonces las ecuaciones del sistema (1)—(5) para $2 \leqslant i \leqslant N-2$ serán satisfechas.

Las relaciones recurrentes (10) conectan α_{i+1}, β_{i+1} y γ_{i+1} con α_i, α_{i-1}, β_i, β_{i-1}, γ_i y γ_{i-1}. Por eso, si son dados α_i, β_i y γ_i para $i = 1, 2$, entonces por las fórmulas (10) se pueden hallar sucesivamente los coeficientes α_i, β_i y γ_i para $3 \leqslant i \leqslant N - 1$.

Hallemos α_i, β_i y γ_i para $i = 1, 2$. De (1) y de la fórmula (6) para $i = 0$ obtenemos inmediatamente

$$\alpha_1 = d_0/c_0, \quad \beta_1 = e_0/c_0, \quad \gamma_1 = f_0/c_0. \tag{11}$$

A continuación, sustituyendo el valor de (8) para $i = 1$ en (2), obtenemos

$$(c_1 - b_1\alpha_1)\, y_1 = (d_1 - b_1\beta_1)\, y_2 - e_1 y_3 + f_1 + b_1\gamma_1.$$

Por consiguiente, (2) será cumplida, si suponemos que

$$\alpha_2 = \frac{d_1 - b_1\beta_1}{c_1 - b_1\alpha_1}, \quad \beta_2 = \frac{e_1}{c_1 - b_1\alpha_1}, \quad \gamma_2 = \frac{f_1 + b_1\gamma_1}{c_1 - b_1\alpha_1}. \tag{12}$$

Así, utilizando (10)—(12), se puede hallar α_i, β_i y γ_i para $1 \leqslant i \leqslant N - 1$. Quedan por determinar α_N, γ_N e y_N, que entran en la fórmula (7).

Para esto utilicemos las ecuaciones (4) y (5). Sustituyendo (8) y (9) para $i = N - 1$ en (4) y comparando la expresión obtenida con (7), hallaremos, que α_N y γ_N se determinan por las fórmulas (10) para $i = N - 1$. Hallemos ahora y_N. Para esto sustituyamos (6) con $i = N - 2$ y (7) en la ecuación (5). Obtenemos

$$[c_N - a_N\beta_{N-1} + \alpha_N (a_N\alpha_{N-1} - b_N)]\, y_N =$$
$$= f_N - a_N\gamma_{N-1} - \gamma_N (a_N\alpha_{N-1} - b_N)$$

o

$$y_N = \gamma_{N+1},$$

donde γ_{N+1} se determina por la fórmula (10) para $i = N$.

Uniendo las fórmulas obtenidas más arriba, escribamos el algoritmo de factorización derecha para el sistema (1)—(5) en la siguiente forma:

1) por las fórmulas

$$\alpha_{i+1} = \frac{1}{\Delta_i}\, [d_i + \beta_i (a_i\alpha_{i-1} - b_i)], \quad i = 2, 3, \ldots, N - 1, \tag{13}$$

$$\alpha_1 = \frac{d_0}{c_0}, \quad \alpha_2 = \frac{1}{\Delta_1}(d_1 - \beta_1 b_1),$$

$$\gamma_{i+1} = \frac{1}{\Delta_i}\, [f_i - a_i\gamma_{i-1} - \gamma_i (a_i\alpha_{i-1} - b_i)], \quad i = 2, 3, \ldots, N,$$

$$\gamma_1 = \frac{f_0}{c_0}, \quad \gamma_2 = \frac{1}{\Delta_1}(f_1 + b_1\gamma_1), \tag{14}$$

$$\beta_{i+1} = e_i/\Delta_i, \quad i = 1, 2, \ldots, N-2, \quad \beta_1 = e_0/c_0, \tag{15}$$

dondo

$$\Delta_i = c_i - a_i\beta_{i-1} + \alpha_i (a_i\alpha_{i-1} - b_i), \ 2 \leqslant i \leqslant N,$$

$$\Delta_1 = c_1 - b_1\alpha_1, \tag{16}$$

se encuentran los coeficiontes de factorización α_i, β_i y γ_i:

2) las incógnitas y_i se encuentran sucesivamonto por las fórmulas

$$y_i = \alpha_{i+1}y_{i+1} - \beta_{i+1}y_{i+2} + \gamma_{i+1}, \ i = N-2, \\ N-3, \ldots, 0, \tag{17}$$

$$y_{N-1} = \alpha_N y_N + \gamma_N, \quad y_N = \gamma_{N+1}.$$

Al algoritmo construido también, lo llamaremos algoritmo de *factorización monótona*.

OBSERVACIÓN. No representa trabajo construir el algoritmo de factorización izquierda y también el algoritmo de factorizaciónes opuostas para el sistoma (1)—(5).

Calculemos el número de operaciones aritméticas para el algoritmo (13)—(17). Para la realización do (13)—(17) se exige: $8N - 5$ operaciones de suma y resta, $8N - 5$ operaciones de multiplicación y $3N$ operaciones de división. Si no hacemos diferencia entre los tiempos de ejecución de las operaciones aritméticas en las CE, entonces el número general do operaciones para el algoritmo propuesto es $Q = 19N - 10$.

2. **Fundamentación del método.** El algoritmo de factorización (13)—(17) construido más arriba se llamará *correcto*, si para cualquier $2 \leqslant i \leqslant N$ es cierta la desigualdad

$$\Delta_i = c_i - a_i\beta_{i-1} + \alpha_i (a_i\alpha_{i-1} - b_i) \neq 0,$$

$$\Delta_1 = c_1 - \alpha_1 b_1 \neq 0.$$

El siguiente lema da una condición suficiente para la corrección del algoritmo (13)—(17).

LEMA 4. *Supongamos que los coeficientes del sistema (1)—(5) satisfacen las condiciones*

$$|a_i| > 0, \ 2 \leqslant i \leqslant N, \ |b_i| > 0, \ 1 \leqslant i \leqslant N,$$

$$|d_i| > 0, 0 \leqslant i \leqslant N-1, |e_i| > 0, 0 \leqslant i \leqslant N-2,$$

y las condiciones

$$|c_0| \geqslant |d_0| + |e_0|, \quad |c_1| \geqslant |b_1| + |d_1| + |e_1|,$$
$$|c_N| \geqslant |a_N| + |b_N|, \quad |c_{N-1}| \geqslant |a_{N-1}| + |b_{N-1}| + |d_{N-1}|, \qquad (18)$$
$$|c_i| \geqslant |a_i| + |b_i| + |d_i| + |e_i|, \quad 2 \leqslant i \leqslant N-2,$$

al mismo tiempo al menos una de las desigualdades (18) se alcanza estrictamente. Entonces el algoritmo (13)—(17) es correcto y además tienen lugar las desigualdades

$$|\alpha_i| + |\beta_i| \leqslant 1, \quad 1 \leqslant i \leqslant N-1, \quad |\alpha_N| \leqslant 1.$$

En efecto, en virtud de las condiciones del lema, de (13) y de (15), obtenemos

$$|\alpha_1| + |\beta_1| = \frac{|d_0| + |e_0|}{|c_0|} \leqslant 1.$$

Seguidamente, empleando la desigualdad obtenida $1 - |\alpha_1| \geqslant |\beta_1|$ hallamos

$$|c_1 - b_1\alpha_1| \geqslant |c_1| - |b_1||\alpha_1| \geqslant |b_1| \times (1 - |\alpha_1|) + |d_1| + |e_1| \geqslant |b_1||\beta_1| + |d_1| + |e_1| \geqslant |d_1 - b_1\beta_1| + |e_1| > 0.$$

De aquí y de (13)—(15) se deduce la estimación

$$|\alpha_2| + |\beta_2| = \frac{|d_1 - \beta_1 b_1| + |e_1|}{|c_1 - b_1\alpha_1|} \leqslant 1.$$

A continuación realizaremos la demostración por inducción. Supongamos que se cumplen las desigualdades

$$|\alpha_{i-1}| + |\beta_{i-1}| \leqslant 1, \quad |\alpha_i| + |\beta_i| \leqslant 1. \qquad (19)$$

Mostremos, que entonces serán válidas las desigualdades

$$\Delta_i = c_i - a_i\beta_{i-1} + \alpha_i(a_i\alpha_{i-1} - b_i) \neq 0,$$
$$|\alpha_{i+1}| + |\beta_{i+1}| \leqslant 1.$$

En efecto, de (18) y (19) obtenemos

$$|\Delta_i| \geqslant |c_i| - |a_i||\beta_{i-1}| - |\alpha_i||\alpha_{i-1}||a_i| - |\alpha_i||b_i| \geqslant |a_i|(1 - |\beta_{i-1}|) + |b_i|(1 - |\alpha_i|) - |\alpha_i||\alpha_{i-1}||a_i| + |d_i| + |e_i| \geqslant |a_i||\alpha_{i-1}| + |b_i||\beta_i| - |\alpha_i||\alpha_{i-1}||a_i| + |d_i| + |e_i| \geqslant$$

$$\geqslant |a_i| |\alpha_{i-1}| (1 - |\alpha_i|) + |d_i - b_i\beta_i| + |e_i| \geqslant$$
$$\geqslant |a_i| |\alpha_{i-1}| |\beta_i| + |d_i - b_i\beta_i| + |e_i| \geqslant$$
$$\geqslant |d_i + \beta_i (a_i\alpha_{i-1} - b_i)| +$$
$$+ |e_i| > 0, \quad i < N - 2. \qquad (20)$$

De aquí y de (13), (15) hallamos

$$|\alpha_{i+1}| + |\beta_{i+1}| = \frac{|d_i + \beta_i (a_i\alpha_{i-1} - b_i)| + |e_i|}{|\Delta_i|} \leqslant 1,$$
$i \leqslant N - 2.$

Así sucesivamente, para $i = N - 1$ tendremos en lugar de (20) la estimación

$$|\Delta_{N-1}| \geqslant |a_{N-1}| |\alpha_{N-2}| |\beta_{N-1}| +$$
$$+ |b_{N-1}| |\beta_{N-1}| + |d_{N-1}| > 0.$$

Además, de aquí obtenemos

$$|\Delta_{N-1}| \geqslant |d_{N-1} + \beta_{N-1} (a_{N-1}\alpha_{N-2} - b_{N-1})|,$$

y, por consiguiente,

$$|\alpha_N| = \frac{1}{|\Delta_{N-1}|} |d_{N-1} + \beta_{N-1} (a_{N-1}\alpha_{N-2} - b_{N-1})| \leqslant 1.$$

Queda por mostrar, que $\Delta_N \neq 0$. Se tiene

$$|\Delta_N| \geqslant |c_N| - |a_N| |\beta_{N-1}| -$$
$$- |\alpha_N| |\alpha_{N-1}| |a_N| - |\alpha_N| |b_N| =$$
$$= |c_N| - |a_N| - |b_N| + |a_N| (1 - |\beta_{N-1}|) +$$
$$+ |b_N| (1 - |\alpha_N|) - |\alpha_N| |\alpha_{N-1}| |a_N| \geqslant$$
$$\geqslant |c_N| - |a_N| - |b_N| + (1 - |\alpha_N|) ((1 - |\beta_{N-1}|) \times$$
$$\times |a_N| + |b_N| (1 - |\alpha_N|).$$

En virtud de las suposiciones del lema es fácil obtener, que al menos en una de las desigualdades $|c_N| \geqslant |a_N| + |b_N|$, $|\alpha_N| \leqslant 1$, se alcanza la desigualdad estricta. De aquí se deduce que $\Delta_N \neq 0$. El lema está demostrado.

OBSERVACIÓN. De las estimaciones $|\alpha_i| + |\beta_i| \leqslant 1$, indicados en el lema 4 se deduce que si durante el cómputo de y_N se admite un error, entonces éste no crecerá en el cálculo por las fórmulas (17).

3. **Variante de factorización no monótona.** Citemos ahora el algoritmo del método de factorización que se obtiene si se busca la solución del sistema (1)-(5) por el método de Gauss con elección del

elemento principal por fila. Este algoritmo será correcto siendo única la condición de no degeneración de la matriz A del sistema (1)-(5). Como el método de construcción del algoritmo es análogo al examinado en el punto 4 del § 2, entonces nos limitaremos a citar la forma definitiva del algoritmo.

1) Se dan los valores iniciales: $C = c_0$, $D = d_0$, $B = b_1$, $Q = c_1$, $S = a_2$, $T = b_2$, $R = 0$, $A = a_3$, $F = f_0$, $\Phi = f_1$, $G = f_2$, $H = f_3$ y se supone $\varkappa_0 = 0$, $\eta_0 = 1$.

2) Sucesivamente para $i = 0, 1, \ldots, N - 2$ en dependencia de la situación se realizan las operaciones descritas en los puntos a), b) ó c):

a) si $|C| \geqslant |D|$ y $|C| \geqslant |e_i|$, entonces

$$\alpha_{i+1} = D/C, \ \beta_{i+1} = e_i/C, \ \gamma_{i+1} = F/C,$$
$$C = Q - B\alpha_{i+1}, \ D = d_{i+1} - B\beta_{i+1}, \ F = \Phi + B\gamma_{i+1},$$
$$B = T - S\alpha_{i+1}, \ Q = c_{i+2} - S\beta_{i+1}, \ \Phi = G - S\gamma_{i+1},$$
$$S = A - R\alpha_{i+1}, \ T = b_{i+3} - R\beta_{i+1}, \ G = H + R\gamma_{i+1} \quad (21)$$
$$\left.\begin{aligned} &R = 0, \ A = a_{i+4}, \ H = {}_{i+4}, \\ &\theta_{i+1} = \varkappa_i, \ \varkappa_{i+1} = \eta_i, \ \eta_{i+1} = i + 2; \end{aligned}\right\} \quad (22)$$

b) si $|D| > |C|$ y $|D| \geqslant |e_i|$, entonces

$$\alpha_{i+1} = C/D, \quad \beta_{i+1} = -e_i/D, \quad \gamma_{i+1} = -F/D,$$
$$C = Q\alpha_{i+1} - B, \quad D = Q\beta_{i+1} + d_{i+1}, \quad F = \Phi - Q\gamma_{i+1},$$
$$B = T\alpha_{i+1} - S, \quad Q = T\beta_{i+1} + c_{i+2}, \quad \Phi = T\gamma_{i+1} + G,$$
$$S = A\alpha_{i+1} - R, \quad T = A\beta_{i+1} + b_{i+3}, \quad G = H - A\gamma_{i+1}, \quad (23)$$
$$\left.\begin{aligned} &R = 0, \ A = a_{i+4}, \ H = f_{i+4}, \\ &\theta_{i+1} = \eta_i, \ \varkappa_{i+1} = \varkappa_i, \ \eta_{i+1} = i + 2; \end{aligned}\right\} \quad (24)$$

c) si $|e_i| > C$ y $|e_i| > |D|$, entonces

$$\alpha_{i+1} = D/e_i, \quad \beta_{i+1} = C/e_i, \quad \gamma_{i+1} = F/e_i,$$
$$C = Q - d_{i+1}\alpha_{i+1}, \quad D = B - d_{i+1}\beta_{i+1}, \quad F = \Phi + d_{i+1}\gamma_{i+1},$$
$$B = T - c_{i+2}\alpha_{i+1}, \quad Q = S - c_{i+2}\beta_{i+1}, \quad \Phi = G - c_{i+2}\gamma_{i+1},$$
$$S = A - b_{i+3}\alpha_{i+1}, \quad T = R - b_{i+3}\beta_{i+1}, \quad G = H + b_{i+3}\gamma_{i+1}, \quad (25)$$
$$\left.\begin{aligned} &R = -a_{i+4}\alpha_{i+1}, \ A = -a_{i+4}\beta_{i+1}, \ H = f_{i+4} - a_{i+4}\gamma_{i+1}, \\ &\theta_{i+1} = i + 2, \ \varkappa_{i+1} = \eta_i, \ \eta_{i+1} = \varkappa_i. \end{aligned}\right\} \quad (26)$$

OBSERVACIÓN. Para $i \geqslant N - 3$ no es necesario realizar los cálculos por las fórmulas (22), (24) o (26) y para $i = N - 2$ no se realizan los cálculos por las fórmulas (21), (23) y (25).

3) Si $|C| \geqslant |D|$, entonces $\alpha_N = D/C$, $\gamma_N = F/C$, $\gamma_{N+1} = (\Phi + B\gamma_N)/(Q - B\alpha_N)$, $\theta_N = \varkappa_{N-1}$ y $\varkappa_N = \eta_{N-1}$. Si $|D| > |C|$, entonces $\alpha_N = C/D$, $\gamma_N = -F/D$, $\gamma_{N+1} = (\Phi - Q\gamma_N)/(Q\alpha_N - B)$, $\theta_N = \eta_{N-1}$ y $\varkappa_N = \varkappa_{N-1}$.

4) Se calculan las incógnitas $y_n = \gamma_{N+1}$, $y_m = \alpha_N y_n + \gamma_N$, $m = \theta_N$, $n = \varkappa_N$, y después sucesivamente para $i = N - 2, N - 3, \ldots, 0$ se determinan las restantes incógnitas $y_m = \alpha_{i+1}y_n - \beta_{i+1}y_k + \gamma_{i+1}$, $m = \theta_{i+1}$, $n = \varkappa_{i+1}$, $k = \eta_{i+1}$.

Examinemos un ejemplo de aplicar el método de factorización no monótona. En el punto 3 del § 3 del cap. 1 fue resuelto el siguiente

problema de contorno de diferencias:

$$\begin{aligned} & y_0 - y_1 + 2y_2 = 2, && i = 0, \\ & -y_0 + y_1 - y_2 + y_3 = 0, && i = 1, \\ & y_{i-2} - y_{i-1} + 2y_i - y_{i+1} + y_{i+2} = 0, && 2 \leqslant i \leqslant N-2, \\ & y_{N-3} - y_{N-2} + y_{N-1} - y_N = 0, && i = N-1, \\ & 2y_{N-2} - y_{N-1} + y_N = 0, && i = N. \end{aligned} \tag{27}$$

Si N es el par y no múltiplo de 3, entonces el sistema (27) tiene la solución única

$$y_i = -\cos\frac{i\pi}{2} - \operatorname{sen}\frac{i\pi}{2}, \quad 0 \leqslant i \leqslant N. \tag{28}$$

Resulta sencillo cerciorarse de que el algoritmo de factorización monótona para el sistema (27) no es correcto — al calcular los coeficientes de factorización α_2, β_2 y γ_2 será necesario dividir por cero. El algoritmo de factorización no monótona permite obtener la solución exacta (28). Mostremos como ilustración de este algoritmo (tabla 2) los valores de los coeficientes de factorización α_i, β_i y γ_i y además θ_i, $\varkappa_i$ y γ_i para $N = 10$.

Tabla 2

i	0	1	2	3	4	5	6	7	8	9	10	11
α_i		1/2	1/2	—1/2	0	0	0	—1/3	—1/3	0	1	
β_i		1/2	1/2	—1/2	1	—1	1	—2/3	—2/3	1		
γ_i		1	1	—1	—2	—2	2	4/3	—2/3	0	—2	1
θ_i		2	3	4	0	5	6	7	9	8	1	
$\varkappa_i$		1	0	1	1	1	1	1	8	1	10	
η_i		0	1	0	5	6	7	8	1	10		
y_i	—1	—1	1	1	—1	—1	1	1	—1	—1	1	

De la tabla se ve, que las incógnitas y_i se determinan en el siguiente orden: $y_{10}, y_1, y_8, y_2, y_7, y_6, y_5, y_0, y_4, y_3, y_2$ es decir en un orden no monótono.

§ 4. Método de factorización matricial

1. Sistema de ecuaciones vectoriales. Más arriba fue señalado que uno de los métodos de solución de problemas de contorno para ecuaciones diferenciales ordinarias de orden superior es la reducción a un sistema de ecuaciones de primer orden con la consecuente aproximación de este sistema por un esquema de diferencias. Como resultado obtenemos un sistema vectorial bipuntual de ecuaciones

con condiciones de contorno de primer género. En forma general este sistema se escribe de la siguiente manera:

$$P_{i+1}V_{i+1} - Q_iV_i = F_{i+1}, \quad 0 \leqslant i \leqslant N-1,$$
$$P_0V_0 = F_0, \quad Q_NV_N = F_{N+1}, \tag{1}$$

donde V_i es el vector de las incógnitas de dimensión M, P_{i+1} y Q_i son las matrices cuadradas $M \times M$ para $0 \leqslant i \leqslant N-1$, P_0 y Q_N son las matrices triangulares de dimensiones $M_1 \times M$ y $M_2 \times M$ respectivamente, siendo $M_1 + M_2 = M$. El vector F_{i+1} para $0 \leqslant i \leqslant N-1$ tiene dimensión M, y los vectores F_0 y F_{N+1} tienen dimensión M_1 y M_2, respectivamente.

Observemos que otro método de solución de las ecuaciones diferenciales indicadas es la aproximación directa de estas ecuaciones por esquemas de diferencias. En este caso obtenemos un sistema de ecuaciones escalares multipuntuales. Los métodos de solución de ecuaciones escalares tripuntuales y pentapuntuales fueron estudiados por nosotros en los §§ 1-3. Si se aproxima un sistema de ecuaciones diferenciales ordinarias de orden superior, entonces aparece un sistema de ecuaciones vectoriales multipuntuales. Sin embargo, tanto los sistemas escalares como los sistemas vectoriales de ecuaciones multipuntuales pueden ser reducidos a sistemas del tipo (1). A su vez, a cualquier método de solución de (1) corresponderá un cierto método de solución del sistema inicial multipuntual. Explicaremos la idea de la transformación indicada mediante el ejemplo del sistema de ecuaciones pentapuntuales, examinado en el § 3 (véase (1)-(5)). Si designamos.

$$Y_i = (y_{i+1}, y_i, y_{i-1}, y_{i-2}), \quad 2 \leqslant i \leqslant N-1,$$

$$F_{i+1} = (f_i, 0, 0, 0), \quad 2 \leqslant i \leqslant N-2,$$

$$F_2 = (f_0, f_1), \quad F_N = (f_{N-1}, f_N),$$

entonces teniendo en cuenta las relaciones idénticas entre Y_{i+1} y Y_i, el sistema señalado del § 3 se escribe en la forma

$$P_{i+1}Y_{i+1} - Q_iY_i = F_{i+1}, \quad 2 \leqslant i \leqslant N-2,$$
$$P_2Y_2 = F_2, \quad Q_{N-1}Y_{N-1} = F_N, \tag{2}$$

donde

$$P_{i+1}=\begin{Vmatrix} e_i & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{Vmatrix},\quad Q_i=\begin{Vmatrix} d_i & -c_i & b_i & -a_i\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & -1 & 0\end{Vmatrix},\quad 2\leqslant i\leqslant N-2,$$

$$P_2=\begin{Vmatrix} 0 & e_0 & -d_0 & c_0\\ e_1 & -d_1 & c_1 & -b_1\end{Vmatrix},$$

$$Q_{N-1}=\begin{Vmatrix} -d_{N-1} & c_{N-1} & -b_{N-1} & a_{N-1}\\ c_N & -b_N & a_N & 0\end{Vmatrix}.$$

En el caso dado, $M_1=M_2=2$ y $M=4$.

A pesar de que las ecuaciones vectoriales multipuntuales se pueden reducir a la forma (1) y limitarse a la construcción de un método de solución solamente del sistema (1), nosotros examinaremos por separado la clase de *las ecuaciones vectoriales tripuntuales*. Además, en el punto 3 reduciremos (1) a un sistema de ecuaciones vectoriales tripuntuales y obtendremos el método de solución del sistema (1) como una variante del método de solución de las ecuaciones tripuntuales.

Antes de describir la forma general de las ecuaciones tripuntuales, examinemos un ejemplo. Mostraremos como un esquema de diferencias para la ecuación elíptica más simple se reduce a un sistema de ecuaciones tripuntuales de un tipo especial.

Supongamos que sobre la red rectangular $\bar{\omega}=\{x_{ij}=(ih_1, jh_2)\in\bar{G},\ 0\leqslant i\leqslant M,\ 0\leqslant j\leqslant N,\ l_1=Mh_1,\ l_2=Nh_2\}$ con frontera γ introducida en el rectángulo $\bar{G}=\{0\leqslant x_\alpha\leqslant l_\alpha,\ \alpha=1, 2\}$, se exige hallar la solución *del problema de Dirichlet de diferencias para la ecuación de Poisson*

$$\begin{aligned} &y_{\bar{x}_1x_1}+y_{\bar{x}_2x_2}=-\varphi(x), && x\in\omega,\\ &y(x)=g(x), && x\in\gamma,\end{aligned}\tag{3}$$

donde

$$y_{\bar{x}_1x_1}=\frac{1}{h_1^2}[y(i+1, j)-2y(i, j)+y(i-1, j)],$$

$$y_{\bar{x}_2x_2}=\frac{1}{h_2^2}[y(i, j+1)-2y(i, j)+y(i, j-1)],$$

$$y(i, j)=y(x_{ij}).$$

Transformemos el esquema (3). Para ello multipliquemos (3) por $(-h_2^2)$ y anotemos por puntos la derivada de diferencias $y_{\bar{x}_2 x_2}$. Cuando $1 \leqslant j \leqslant N-1$, tendremos:

para $2 \leqslant i \leqslant M-2$

$$-y(i, j-1) + [2y(i, j) - h_2^2 y_{\bar{x}_1 x_1}(i, j)] - \\ -y(i, j+1) = h_2^2 \varphi(i, j);$$

para $i = 1$

$$-y(i, j-1) + \left[2y(i, j) - \frac{h_2^2}{h_1^2}(y(i+1, j) - 2y(i, j))\right] - \\ -y(i, j+1) = h_2^2 \bar{\varphi}(i, j);$$

para $i = M-1$

$$-y(i, j-1) + \left[2y(i, j) - \frac{h_2^2}{h_1^2}(y(i-1, j) - 2y(i, j))\right] - \\ -y(i, j+1) = h_2^2 \bar{\varphi}(i, j),$$

donde

$$\bar{\varphi}(1, j) = \varphi(1, j) + \frac{1}{h_1^2} g(0, j),$$

$$\bar{\varphi}(M-1, j) = \varphi(M-1, j) + \frac{1}{h_1^2} g(M, j).$$

Además, para $j = 0, N$ tenemos

$$y(i, 0) = g(i, 0), \; y(i, N) = g(i, N), \; 1 \leqslant i \leqslant M-1$$

Denotemos ahora mediante Y_j el vector de dimensión $M-1$ cuyas componentes son los valores de la función reticular $y(i, j)$ en los nodos interiores de la red $\bar{\omega}$ sobre la j-ésima fila:

$$Y_j = (y(1, j), y(2, j), \ldots, y(M-1, j)), \; 0 \leqslant j \leqslant N$$

y mediante F_j el vector de dimensión $M-1$

$$F_j = (h_2^2 \bar{\varphi}(1, j), h_2^2 \varphi(2, j), \ldots, h_2^2 \varphi(M-2, j), \\ h_2^2 \bar{\varphi}(M-1, j)), \quad 1 \leqslant j \leqslant N-1,$$

$$F_j = (g(1, j), g(2, j), \ldots, g(M-1, j)), \; j = 0, N.$$

Definamos también la matriz cuadrada C con dimensión $(M-1)(M-1)$ de la siguiente forma:

$$CV = (\Lambda v(1), \Lambda v(2), \ldots, \Lambda v(M-1)),$$

$$V = (v(1), v(2), \ldots, v(M-1)),$$

donde el operador de diferencias Λ es

$$\Lambda v(i) = 2v(i) - h_2^2 v_{\bar{x}_1 x_1}(i), \quad 1 \leqslant i \leqslant M-1,$$
$$v(0) = v(M) = 0.$$

Es fácil ver, que C es la matriz tridiagonal del tipo

$$C = \begin{Vmatrix} 2(1+\alpha) & -\alpha & 0 & \ldots & 0 & 0 & 0 \\ -\alpha & 2(1+\alpha) & -\alpha & \ldots & 0 & 0 & 0 \\ 0 & -\alpha & 2(1+\alpha) & \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & 2(1+\alpha) & -\alpha & 0 \\ 0 & 0 & 0 & \ldots & -\alpha & 2(1+\alpha) & -\alpha \\ 0 & 0 & 0 & \ldots & 0 & -\alpha & 2(1+\alpha) \end{Vmatrix}, \qquad (4)$$

donde $\alpha = h_2^2 h_1^2$, además C es la matriz con predominancia diagonal, ya que $|1 + \alpha| > |\alpha|$, $\alpha > 0$, y por consiguiente, no es degenerada.

Utilizando las notaciones introducidas, las relaciones obtenidas más arriba se pueden escribir en la forma del siguiente sistema de ecuaciones vectoriales tripuntuales:

$$\begin{aligned} &-Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1, \\ &Y_0 = F_0, \quad Y_N = F_N. \end{aligned} \qquad (5)$$

Este es el sistema tripuntual buscado de tipo especial con coeficientes constantes.

El problema (5) es un caso especial del siguiente problema general: hallar los vectores Y_j $(0 \leqslant j \leqslant N)$, que satisfacen el siguiente sistema:

$$\begin{aligned} &C_0 Y_0 - B_0 Y_1 = F_0, && j = 0, \\ &-A_j Y_{j-1} + C_j Y_j - B_j Y_{j+1} = F_j, && 1 \leqslant j \leqslant N-1, \\ &-A_N Y_{N-1} + C_N Y_N = F_N, && j = N, \end{aligned} \qquad (6)$$

donde Y_j y F_j son los vectores de dimensión M_j, C_j es la matriz cuadrada $M_j \times M_j$, A_j y B_j son las matrices rectangulares de dimensiones $M_j \times M_{j-1}$ y $M_j \times M_{j+1}$ respectivamente.

A los sistemas del tipo (6) se reducen los esquemas de diferencias para las ecuaciones elípticas de segundo orden con coeficientes variables en la región arbitraria de cualquier cantidad de mediciones. En el caso bidimensional, como en el ejemplo analizado más arriba, al vector Y_j lo forman las incógnitas sobre la j-ésima fila de la red $\bar{\omega}$, y en el caso de tres dimensiones las incógnitas sobre la

j-ésima capa de la red $\bar{\omega}$. En el último caso, C_j es la matriz tridiagonal por bloques con matrices triagonales sobre la diagonal principal.

Para la solución del sistema (6) examinaremos *el método de factorización matricial*, el cual es análogo al método de factorización para ecuaciones escalares tripuntuales.

2. **Factorización para ecuaciones vectoriales tripuntuales.** Construyamos el método de resolución del sistema de ecuaciones vectoriales tripuntuales (6). Este sistema es cercano a un sistema de ecuaciones tripuntuales escalares, cuyo método de solución fue estudiado por nosotros en el § 1. Por eso, buscaremos la solución del problema (6) en la forma

$$Y_j = \alpha_{j+1} Y_{j+1} + \beta_{j+1}, \quad j = N-1, N-2, \ldots, 0, \tag{7}$$

donde α_{j+1} es, por ahora, una matriz rectangular indeterminada de dimensiones $M_j \times M_{j+1}$, y β_{j+1} es el vector de dimensión M_j. De la fórmula (7) y de las ecuaciones del sistema (6) para $1 \leqslant j \leqslant N-1$ se encuentran (como en el caso de factorización común) las relaciones recurrentes para calcular la matriz α_j y los vectores β_j. De (7) y de las ecuaciones (6) para $j = 0, N$, se hallan los valores iniciales α_1, β_1 y Y_N, que permiten comenzar el cálculo según las relaciones recurrentes. Escribamos las fórmulas definitivas del algoritmo del método propuesto, al cual llamaremos *método de factorización matricial*:

$$\alpha_{j+1} = (C_j - A_j\alpha_j)^{-1} B_j,$$
$$j = 1, 2, \ldots, N-1, \quad \alpha_1 = C_0^{-1} B_0, \tag{8}$$

$$\beta_{j+1} = (C_j - A_j\alpha_j)^{-1} (F_j + A_j\beta_j),$$
$$j = 1, 2, \ldots, N, \quad \beta_1 = C_0^{-1} F_0, \tag{9}$$

$$Y_j = \alpha_{j+1} Y_{j+1} + \beta_{j+1},$$
$$j = N-1, N-2, \ldots, 0, \quad Y_N = \beta_{N+1}. \tag{10}$$

Diremos que el algoritmo (8)-(10) es *correcto*, si las matrices C_0 y $C_j - A_j\alpha_j$ para $1 \leqslant j \leqslant N$ no son degenerados. Antes de dar la definición de estabilidad del algoritmo (8)-(10) recordemos algunos aspectos del álgebra lineal.

Sea A una $m \times n$ matriz rectangular arbitraria. Sea $\| x \|_n$ la norma del vector x en el espacio n-dimensional H_n.

Entonces la norma de A se define por la igualdad

$$\| A \| = \sup_{x \neq 0} \frac{\| Ax \|_m}{\| x \|_n} .$$

Es evidente, que la norma A se define por la matriz A y por las normas vectoriales, que han sido introducidas en H_n y H_m. Para el caso de las normas euclidianas en H_n y H_m ($\| x \|_n^2 = \sum_{i=1}^{n} x_i^2$) tenemos $\| A \| = \sqrt{\rho}$, donde ρ es el valor propio de módulo máximo de la matriz A^*A.

De la definición de la norma se deduce la relación evidente $\| Ax \|_m \leqslant \| A \| \| x \|_n$.

Así sucesivamente, sean dadas las matrices A y B de dimensiones $m \times n$ y $n \times k$ respectivamente. Introduciendo normas vectoriales en H_m, H_k y H_n y definiendo con su ayuda las normas de las matrices A, B y AB, obtenemos la desigualdad $\| AB \| \leqslant \| A \| \| B \|$.

Diremos que el algoritmo es *estable*, si se cumple la estimación $\| \alpha_j \| \leqslant 1$ para $1 \leqslant j \leqslant N$, (se supone, que en los espacios de dimensión finita H_{M_j}, a los cuales pertenecen los vectores Y_j, se han introducido norma de un mismo tipo, por ejemplo euclidiana).

LEMA 5. *Si C_j para $0 \leqslant j \leqslant N$ son las matrices no degeneradas*, y A_j y B_j *son las matrices no nulas para* $1 \leqslant j \leqslant N - 1$ *y se cumplen las condiciones*

$$\| C_0^{-1} B_0 \| \leqslant 1, \quad \| C_N^{-1} A_N \| \leqslant 1,$$

$$\| C_j^{-1} A_j \| + \| C_j^{-1} B_j \| \leqslant 1. \quad 1 \leqslant j \leqslant N - 1,$$

y si además al menos en una de las desigualdades tiene lugar la desigualdad estricta, entonces el algoritmo del método de factorización matricial (8)-(10) es estable y correcto.

Citemos solamente la etapa fundamental, dejando al lector la culminación de la demostración del lema. La demostración utiliza una afirmación conocida: si para la matriz cuadrada S ocurre la estimación $\| S \| \leqslant q \leqslant 1$, entonces existe la matriz inversa a $E - S$, con todo $\| (E - S)^{-1} \| \leqslant 1/(1 - q)$.

Supongamos ahora que $\| \alpha_j \| \leqslant 1$. De aquí y de las condiciones del lema tendremos

$$\| C_j^{-1} A_j \alpha_j \| \leqslant \| C_j^{-1} A_j \| \leqslant 1 - \| C_j^{-1} B_j \| < 1 .$$

Puesto que $C_j^{-1} A_j \alpha_j$ es la matriz cuadrada, entonces, por lo tanto, existen las matrices inversas a la $E - C_j^{-1} A_j^{-1} \alpha_j$

y a la $C_j - A_j\alpha_j$, y además $\| (E - C_j^{-1}A_j\alpha_j)^{-1} \| \leqslant \leqslant 1/\| C_j^{-1}B_j \|$. De aquí y de (8) obtenemos inmediatamente

$$\| \alpha_{j+1} \| \leqslant (E - C_j^{-1}A_j\alpha_j)^{-1} C_j^{-1}B_j \| \leqslant$$
$$\leqslant \| (E - C_j^{-1}A_j\alpha_j)^{-1} \| \| C_j^{-1}B_j \| \leqslant 1.$$

La demostración se culmina por inducción.

Apliquemos el lema 5 al sistema de ecuaciones vectoriales tripuntuales (5), obtenidas del problema de Dirichlet de diferencias para la ecuación de Poisson en un rectángulo. El sistema (5) es un caso particular de (6), donde $C_j = C$, $B_j = A_j = E$, $1 \leqslant j \leqslant N - 1$, $C_0 = C_N = E$, $B_0 = A_N = 0$, y la matriz cuadrada C está dada en (4). Las condiciones del lema 5 para el ejemplo examinado toman la forma $\| C^{-1} \| \leqslant 0{,}5$. Para el caso de la norma euclidiana, en virtud de la simetría de C tenemos

$$\| C^{-1} \| = \max_{h} | \lambda_k (C^{-1}) | = \frac{1}{\min\limits_{k} | \lambda_k (C) |},$$

donde $\lambda_k (C)$ es el valor propio de la matriz C. De la definición de C obtenemos que $\lambda_k (C)$ es el valor propio del operador Λ definido más arriba,

$$\Lambda v (i) - \lambda_k v (i) = (2 - \lambda_k) v (i) - h_2^2 v_{\bar{x}_1 x_1} (i) = 0,$$
$$v (0) = v (M) = 0, \quad 1 \leqslant i \leqslant M - 1.$$

Este problema se reduce mediante el cambio $\lambda_k = 2 + h_2^2 \mu_k$ al problema en valores propios examinado en el punto 1 del § 5 del cap. I, para el operador de diferencias más simple: $v_{\bar{x}_1 x_1} + \mu_k v = 0$, $1 \leqslant i \leqslant M - 1$, $v (0) = v (M) = 0$. Como este problema tiene solución igual a

$$\mu_k = \frac{4}{h_1^2} \operatorname{sen}^2 \frac{k\pi h_1}{2l_1} > 0, \quad k = 1, 2, \ldots, M - 1,$$

entonces $\lambda_k = \lambda_k (C) = 2 + h_2^2 \mu_k > 2$. Por lo tanto, la condición $\| C^{-1} \| \leqslant 0{,}5$ es cumplida. El algoritmo (8)-(10) es correcto y estable al ser aplicado a la solución del sistema (5).

Examinemos ahora el problema sobre el volumen de información intermedia a recordar y sobre la estimación del número de operaciones aritméticas para el algoritmo (8)-(10), considerando para simplificar, que en el sistema (6) todas las matrices son cuadradas y tienen tamaño $M \times M$, y todos los vectores Y_j y F_j tienen dimensión M. En este caso los coeficientes de factorización α_j serán las matrices cuadradas de tamaño $M \times M$ y los vectores β_j tendrán dimensión M.

Para la realización de (8)-(10) hay que conservar todas las matrices α_j para $1 \leqslant j \leqslant N$, todos los vectores β_j para $1 \leqslant j \leqslant N+1$ y la matriz $(C_N - A_N\alpha_N)^{-1}$, que se utiliza para calcular β_{N+1}. Los vectores β_j pueden ser distribuídos en el lugar asignado para los vectores de las incógnitas Y_{j-1}. Para conservar todas las matrices α_j y la matriz $(C_N - A_N\alpha_N)^{-1}$ es necesario recordar $M^2(N+1)$ elementos de estas matrices, ya que en el caso general las matrices α_j son completas y no simétricas. En este caso el volumen de información complementaria recordada es M veces superior al número de incógnitas en el problema, el cual es igual a $M(N+1)$.

Estimemos ahora el número de operaciones aritméticas para el algoritmo (8)-(10), teniendo en cuenta que para la resolución de la serie de problemas (6) con diferentes miembros derechos F_j, las matrices de factorización α_j y la matriz $(C_N - A_N\alpha_N)^{-1}$ pueden ser calculadas solamente una vez, mientras que los vectores β_j y Y_j se vuelven a calcular para cada nuevo problema.

En el caso general las matrices $C_j - A_j\alpha_j$ son completas para cada j. Por eso para su inversión se exigen $O(M^3)$ operaciones aritméticas. A continuación, el producto de $(C_j - A_j\alpha_j)^{-1}$ por la matriz B_j exige no más de $O(M^3)$ operaciones aritméticas. Por eso para el cálculo de α_{j+1} para α_j dado, por medio de la fórmula (8), se necesitan $O(M^3)$ operaciones aritméticas. Para calcular todas las α_j y la matriz $(C_N - A_N\alpha_N)^{-1}$ se exigen $O(M^3N)$ operaciones.

Si la matriz A_j es completa, entonces en el cálculo de β_{j+1} siendo prefijado β_j y calculado $(C_j - A_j\alpha_j)^{-1}$ se exigen $2M^2$ operaciones de multiplicación y $2M^2 - M$ operaciones de suma. Si A_j es la matriz diagonal entonces este número se reduce, exigiéndose $M^2 + M$ multiplicaciones y $2M^2 - M$ sumas. Por consiguiente, en el cálculo de β_j para $2 \leqslant j \leqslant N+1$ se exige en el caso general $2M^2N$ multiplicaciones y $(2M^2 - M)N$ sumas. Añadiendo aquí las operaciones gastadas en el cálculo de β_1 con C_0^{-1} dado (M^2 multiplicaciones y $M^2 - M$ sumas), obtenemos definitivamente $M^2(2N+1)$ multiplicaciones y $M^2(2N+1) - M(N+1)$ sumas.

Para encontrar todos los Y_j para $0 \leqslant j \leqslant N-1$ siendo Y_N dado, se exigen M^2N multiplicaciones y M^2N sumas. De esta forma, para el cálculo de β_j y Y_j se requieren $M^2(3N+1)$ operaciones de multiplicación y $M^2(3N+1) - M(N+1)$ operaciones de sumas. Si no hacemos diferencia entre estas operaciones, entonces esto constituye $Q \approx 6M^2N$ operaciones. Precisamente es necesario gastar tal cantidad de operaciones aritméticas para hallar la solución de cada nuevo problema de la serie. Para resolver un solo problema del tipo (6), cuando es preciso calcular las matrices de factorización α_j, se exigen $Q = O(M^3N + M^2N)$ operaciones.

Supongamos que la serie consiste de n problemas del tipo (6). Entonces se requiere gastar $Q_n = O(M^3N) + 6nM^2N$ operaciones. En este caso el número total de incógnitas en la serie es igual a $nM(N+1)$. De aquí se deduce, que para encontrar una incógnita se necesitan $q \approx 0\left(\frac{M^2}{n}\right) + 6M$ operaciones aritméticas. De esta forma, al aumentar n, el número relativo de operaciones para una incógnita disminuye, sin embargo siempre es mayor que $6M$. En esto se encierra la diferencia esencial entre el método de factorización matricial y el método de factorización escalar, donde el número relativo de operaciones para una incógnita es un número finito que no depende del número de incógnitas.

3. Factorización para ecuaciones vectoiales bipuntuales. Examinemos ahora el método de solución de las ecuaciones vectoriales bipuntuales

$$P_{i+1}V_{i+1} - Q_iV_i = F_{i+1}, \quad 0 \leqslant i \leqslant N-1,$$
$$P_0V_0 = F_0, \quad Q_NV_N = F_{N+1}, \tag{11}$$

donde V_i es el vector de dimensión M, P_{i+1} y Q_i son las matrices cuadradas $M \times M$ para $0 \leqslant i \leqslant N-1$, P_0 y Q_N son las matrices rectangulares de dimensiones $M_1 \times M$ y $M_2 \times M$ respectivamente, siendo $M_1 + M_2 = M$. El vector F_{i+1} para $0 \leqslant i \leqslant N-1$ tiene dimensión M, F_0 y F_{N+1} tienen dimensión M_1 y M_2 respectivamente.

Primeramente reduciremos el sistema (11) a la forma (6). Para esto, representemos las matrices que entran en (11), de la siguiente forma:

$$P_0 = \| P_0^{11} \mid -P_0^{12} \|, \quad Q_N = \| -Q_N^{21} \mid Q_N^{22} \|,$$
$$P_{i+1} = \left\| \frac{P_{i+1}^{11} \; -P_{i+1}^{12}}{P_{i+1}^{21} \; -P_{i+1}^{22}} \right\|, \quad Q_i = \left\| \frac{Q_i^{11} \; -Q_i^{12}}{Q_i^{21} \mid +Q_i^{22}} \right\|, \tag{12}$$

donde P_i^{kl} y Q_i^{kl} para $0 \leqslant i \leqslant N$ son las matrices de dimensiones $M_k \times M_l$, $k, l = 1, 2$. En correspondencia con la representación (12) pongamos

$$V_i = \begin{pmatrix} v_i^1 \\ v_i^2 \end{pmatrix}, \quad 0 \leqslant i \leqslant N, \quad F_{i+1} = \begin{pmatrix} f_{i+1}^1 \\ f_{i+1}^2 \end{pmatrix},$$
$$0 \leqslant i \leqslant N-1, \tag{13}$$
$$F_0 = f_0^1, \quad F_{N+1} = f_{N+1},$$

donde v_i^k y f_i^k son los vectores de dimensión M_k, $k = 1, 2$. Empleando (12) y (13), escribamos el sistema (11) en la siguiente forma:

$$P_0^{11}v_0^1 - P_0^{12}v_0^2 = f_0^1,$$
$$\left.\begin{aligned} &-Q_i^{11}v_i^1 + Q_i^{12}v_i^2 + P_{i+1}^{11}v_{i+1}^1 - \\ &-P_{i+1}^{12}v_{i+1}^2 = f_{i+1}^1, \\ &-Q_i^{21}v_i^1 + Q_i^{22}v_i^2 + P_{i+1}^{21}v_{i+1}^1 - \\ &-P_{i+1}^{22}v_{i+1}^2 = f_{i+1}^2, \end{aligned}\right\} 0 \leqslant i \leqslant N-1, \tag{14}$$
$$-Q_N^{21}v_N^1 + Q_N^{22}v_N^2 = f_{N+1}^2.$$

Introduzcamos ahora nuevos vectores de las incógnitas, poniendo

$$Y_0 = v_0^1, \quad Y_{N+1} = v_N^2, \quad Y_{i+1} = \begin{pmatrix} v_i^2 \\ v_{i+1}^1 \end{pmatrix}, \quad 0 \leqslant i \leqslant N-1,$$

y las matrices,

$$C_0 = P_0^{11}, \quad B_0 = \| P_0^{12} \mid 0^{11} \|, \quad C_{N+1} = Q_N^{22}, \quad A_{N+1} = \| 0^{22} \mid Q_N^{21} \|$$

$$A_1 = \left\| \frac{Q_0^{11}}{Q_0^{21}} \right\|, \quad B_N = \left\| \frac{P_N^{12}}{P_N^{22}} \right\|, \quad A_{i+1} = \left\| \frac{0^{12} \mid Q_i^{11}}{0^{22} \mid Q_i^{21}} \right\|,$$

$1 \leqslant i \leqslant N-1,$

$$B_{i+1} = \left\| \frac{P_{i+1}^{12} \mid 0^{11}}{P_{i+1}^{22} \mid 0^{21}} \right\|, \quad 0 \leqslant i \leqslant N-2,$$

$$C_{i+1} = \left\| \frac{Q_i^{12} \mid P_{i+1}^{11}}{Q_i^{22} \mid P_{i+1}^{21}} \right\|, \quad 0 \leqslant i \leqslant N-1$$

donde 0^{kl} es la matriz nula de dimensiones $M_k \times M_l$, $k, l = 1, 2$.

En estas notaciones el sistema (14) tendrá la forma

$$\begin{aligned} &C_0 Y_0 - B_0 Y_1 = F_0, \quad i = 0, \\ &-A_i Y_{i-1} + C_i Y_i - B_i Y_{i+1} = F_i, \quad 1 \leqslant i \leqslant N, \\ &-A_{N+1} Y_N + C_{N+1} Y_{N+1} = F_{N+1}, \quad i = N+1. \end{aligned} \tag{15}$$

Así, el sistema de ecuaciones vectoriales bipuntuales (11) ha sido reducido a un sistema de ecuaciones vectoriales tripuntuales del tipo (15), para el cual el método de factorización matricial fue construido en el punto 2. Para (15) el algoritmo de factorización matricial tiene la siguiente forma:

$$\alpha_{i+1} = (C_i - A_i \alpha_i)^{-1} B_i, \qquad i = 1, 2, \ldots, N, \quad \alpha_1 = C^{-1} B_0, \tag{16}$$

$$\beta_{i+1} = (C_i - A_i \alpha_i)^{-1} (F_i + A_i \beta_i), \qquad i = 1, 2, \ldots, N+1, \quad \beta_1 = C_0^{-1} F_0 \tag{17}$$

$$Y_i = \alpha_{i+1} Y_{i+1} + \beta_{i+1}, \qquad i = N, N-1, \ldots, 0, \quad Y_{N+1} = \beta_{N+1}, \tag{18}$$

pero las matrices α_1 y α_{N+1} tienen tamaño $M_1 \times M$ y $M \times M_2$ respectivamente, y α_i para $2 \leqslant i \leqslant N$ son las matrices cuadradas de tamaño $M \times M$. Los vectores β_i para $2 \leqslant i \leqslant N+1$ tienen dimensión M y β_1 y β_{N+2} tienen dimensión M_1 y M_2.

Transformemos las fórmulas (16)-(18). Teniendo en cuenta la estructura de las matrices B_i, encontramos, que

las matrices α_i tienen la forma

$$\alpha_1=\|\alpha_1^{12}\mid 0^{11}\|,\qquad \alpha_{N+1}=\left\|\frac{\alpha_{N+1}^{22}}{\alpha_{N+1}^{12}}\right\|,$$

$$\alpha_i=\left\|\frac{\alpha_i^{22}\mid 0^{21}}{\alpha_i^{12}\mid 0^{11}}\right\|,\quad 2\leqslant i\leqslant N. \tag{19}$$

Sustituyendo (19) en (16) y teniendo en cuenta la definición de las matrices A_i, B_i y C_i, obtenemos las fórmulas para calcular α_i^{12} y α_i^{22}

$$\left\|\frac{\alpha_{i+1}^{22}}{\alpha_{i+1}^{12}}\right\|=\left\|\frac{Q_{i-1}^{12}-Q_{i-1}^{11}\alpha_i^{12}\mid P_i^{11}}{Q_{i+1}^{22}-Q_{i-1}^{21}\alpha_i^{12}\mid P_i^{21}}\right\|^{-1}\left\|\frac{P_i^{12}}{P_i^{22}}\right\|,\quad 1\leqslant i\leqslant N, \tag{20}$$

donde $\alpha_1^{12}=(P_0^{11})^{-1}P_0^{12}$. Seguidamente, representando el vector β_i en la forma

$$\beta_1=\beta_1^1,\quad \beta_{N+2}=\beta_{N+2}^2,\quad \beta_i=\binom{\beta_i^2}{\beta_i^1},\quad 2\leqslant i\leqslant N+1 \tag{21}$$

y sustituyendo esta expresión en (17), obtendremos

$$\binom{\beta_{i+1}^2}{\beta_{i+1}^1}=\left\|\frac{Q_{i-1}^{12}-Q_{i-1}^{11}\alpha_i^{12}\mid P_i^{11}}{Q_{i-1}^{22}-Q_{i-1}^{21}\alpha_i^{12}\mid P_i^{21}}\right\|^{-1}\binom{f_i^1+Q_{i-1}^{11}\beta_i^1}{f_i^2+Q_{i-1}^{21}\beta_i^1},\quad 1\leqslant i\leqslant N. \tag{22}$$

$$\beta_{N+2}^2=\|Q_N^{22}-Q_N^{21}\alpha_{N+1}^{12}\|^{-1}(f_{N+1}^2+Q_N^{21}\beta_{N+1}^1), \tag{23}$$

donde $\beta_1^1=\|P_0^{11}\|^{-1}f_0^1$.

Sustituyamos ahora (19) y (21) en (18) y utilicemos las notaciones introducidas para Y_i. Como resultado obtendremos las siguientes fórmulas para el cálculo de las componentes del vector de las incógnitas:

$$\begin{aligned} v_{i-1}^2&=\alpha_{i+1}^{22}v_i^2+\beta_{i+1}^2,\quad i=N,\ N-1,\ \ldots,\ 1,\\ v_N^2&=\beta_{N+2}^2,\\ v_i^1&=\alpha_{i+1}^{12}v_i^2+\beta_{i+1}^1,\quad i=N,\ N-1,\ \ldots,\ 0.\end{aligned} \tag{24}$$

Así, el algoritmo *del método de factorización matricial* para el sistema de ecuaciones vectoriales bipuntuales (11) se describe por las fórmulas (20), (22)-(24).

Ya que estas fórmulas son consecuencia del algoritmo de factorización para la resolución del sistema (15), al cual nosotros redujimos el sistema inicial de ecuaciones vectoriales bipuntuales (11), entonces las condiciones sufi-

cientes de corrección y estabilidad del algoritmo obtenido están formuladas en el lema 5, donde es necesario cambiar N por $N + 1$, y las matrices C_i, A_i y B_i son definidas más arriba.

Utilizando el algoritmo de factorizaciones opuestas para el sistema (15), se puede construir el algoritmo correspondiente para el sistema inicial de ecuaciones vectoriales bipuntuales (11).

4. **Factorización ortogonal para ecuaciones vectoriales bipuntuales.** Examinemos otro método de resolución del sistema de ecuaciones bipuntuales (11), conocido bajo el nombre de *método de factorización ortogonal*. Este método contiene la inversión de las matrices P_i para $1 \leqslant i \leqslant N$ y la ortogonalización de las matrices rectangulares auxiliares.

Buscaremos la solución del sistema (11) en la siguiente forma:

$$V_i = B_i\beta_i + Y_i, \quad 0 \leqslant i \leqslant N, \tag{25}$$

donde B_i, para cualquier i, es la matriz rectangular de tamaño $M \times M_2$ y β_i y Y_i son los vectores de dimensión M_2 y M, respectivamente.

Definiendo B_0 y Y_0 de la condición $P_0B_0 = 0^{12}$, $P_0Y_0 = F_0$, donde 0^{12} es la matriz nula de tamaño $M_1 \times M_2$, obtendremos que V_0 satisface la condición $P_0V_0 = F_0$. Hallemos ahora las fórmulas recurrentes para la sucesiva construcción de las matrices B_i y los vectores Y_i, partiendo de B_0 y Y_0.

Sustituyamos (25) en (11). Si P_{i+1} son las matrices no degeneradas, entonces tendremos

$$B_{i+1}\beta_{i+1} + Y_{i+1} - P_{i+1}^{-1}Q_iB_i\beta_i = P_{i+1}^{-1}(F_{i+1}^{-1} + Q_iY_i)$$
$$0 \leqslant i \leqslant N - 1,$$

ó

$$B_{i+1}\beta_{i+1} + Y_{i+1} - A_{i+1}\beta_i = X_{i+1},$$
$$0 \leqslant i \leqslant N - 1, \tag{26}$$

donde $A_{i+1} = P_{i+1}^{-1}Q_iB_i$, $X_{i+1} = P_{i+1}^{-1}(F_{i+1} + Q_iY_i)$. La matriz A_{i+1} tiene tamaño $M \times M_2$, y el vector X_{i+1} tiene dimensión M.

Definamos B_{i+1} y Y_{i+1} de la siguiente forma:

$$A_{i+1} = B_{i+1}\Omega_{i+1}, \quad Y_{i+1} = X_{i+1} - B_{i+1}\varphi_{i+1}. \tag{27}$$

donde Ω_{i+1} y φ_{i+1} son la matriz cuadrada $M_1 \times M_2$ y el vector de dimensión M_2, no determinados por ahora. Sustituyendo (27) en (26), obtendremos la relación $B_{i+1} \times \times (\beta_{i+1} - \Omega_{i+1}\beta_i) = B_{i+1}\varphi_{i+1}$, la cual se transforma en identidad si ponemos

$$\Omega_{i+1}\beta_i = \beta_{i+1} - \varphi_{i+1}, \quad 0 \leqslant i \leqslant N - 1. \tag{28}$$

Por tanto, si se dan matrices no degeneradas Ω_i para $1 \leqslant \leqslant i \leqslant N$ y vectores φ_i para esos mismos i, entonces por las fórmulas (27) se pueden hallar todas las matrices necesarias B_i y los vectores Y_i para $1 \leqslant i \leqslant N$, partiendo de B_0 y Y_0.

Quedan por definir los vectores β_i. De (25) para $i = N$ y del sistema (11) obtenemos dos relaciones $V_N = B_N\beta_N + + Y_N$, $Q_NV_N = F_{N+1}$ con las incógnitas B_N y Y_N. De aquí para β_N hallamos la ecuación $Q_NB_N\beta_N = F_{N+1} - - Q_NY$ con la matriz cuadrada Q_NB_N de tamaño $M_2 \times M_2$ Esta relación se puede escribir en la forma (28)

$$\Omega_{N+1}\beta_N = \beta_{N+1} - \varphi_{N+1}, \tag{29}$$

donde $\beta_{N+1} = F_{N+1}$, $\varphi_{N+1} = Q_NY_N$, $\Omega_{N+1} = Q_NB_N$.

Si la matriz Ω_{N+1} no es degenerada, entonces según las fórmulas (28) y (29) sucesivamente, comenzando por β_{N+1}, hallamos todos los β_i para $0 \leqslant i \leqslant N$. La solución del sistema (11) puede ser entonces hallada por las fórmulas (25).

Como se tiene arbitrariedad en la elección de las matrices Ω_i y de los vectores φ_i, entonces las fórmulas citadas más arriba describen más bien el principio para construir un método de resolución del sistema (11), que un algoritmo concreto. La elección de Ω_i y φ_i determinados genera un cierto método para el sistema (11). Tales métodos los llamaremos de factorización, en cuyo paso directo se calculan B_i y Y_i, y en el inverso β_i y la solución Y_i.

Detengámonos ahora en un procedimiento para elegir Ω_i y φ_i. Ya que las fórmulas (27) y (28) presuponen la inversión de la matriz Ω_{i+1}, entonces ella debe ser suficientemente fácil de invertir.

En el método examinado de factorización ortogonal la matriz Ω_{i+1} y el vector φ_{i+1} se generan por las exigencias: 1) la matriz B_{i+1} se construye mediante la ortonormalización de las columnas de la matriz A_{i+1}; 2) el vector Y_{i+1} debe ser ortogonal a las columnas de la matriz B_{i+1}.

Consecuencia de estas exigencias son las igualdades

$$B^*_{i+1}B_{i+1} = E^{22}, \quad B^*_{i+1}Y_{i+1} = 0, \tag{29'}$$

donde B^*_{i+1} es la matriz transpuesta a B_{i+1}, y E^{22} es la matriz unidad de tamaño $M_2 \times M_2$.

Hallemos primeramente la expresión para φ_{i+1}. De (27) y (29') obtenemos $0 = B^*_{i+1}Y_{i+1} = B^*_{i+1}X_{i+1} - B^*_{i+1}B_{i+1} \times \times \varphi_{i+1} = B^*_{i+1}X_{i+1} - \varphi_{i+1}$. De esta forma, está determinado el vector φ_{i+1}: $\varphi_{i+1} = B^*_{i+1}X_{i+1}$.

Costruyamos ahora las matrices Ω_{i+1} y B_{i+1}. Existen varios métodos para ortonormalizar las columnas de la matriz A_{i+1}. Nosotros examinaremos el método de Gram—Shmidt.

Supongamos que la matriz A_{i+1} posee rango M_2. Designemos mediante a_k y b_k las k-ésimas columnas de las matrices A_{i+1} y B_{i+1} respectivamente, y por (,) el producto escalar de vectores. En calidad de b_1 tomemos la columna a_1 normalizada

$$b_1 = a_1/\omega_{11}, \quad \omega_{11} = \sqrt{(a_1, a_1)}. \tag{30}$$

Así sucesivamente buscaremos la columna b_k en la forma

$$b_k = \frac{1}{\omega_{kk}}\left(a_k - \sum_{n=1}^{k-1} \omega_{nk} b_n\right), \quad 2 \leqslant k \leqslant M_2, \tag{31}$$

donde los coeficientes ω_{nk} se encuentran de la condición de ortogonalidad del vector b_k a los vectores $b_1, b_2, \ldots$ $\ldots, b_{k-1}$, y ω_{kk} de la condición de normalización de b_k:

$$\omega_{nk} = (b_n, a_k), \quad n = 1, 2, \ldots, k-1,$$

$$\omega_{kk} = \sqrt{(a_k, a_k) - \sum_{n=1}^{k-1} \omega^2_{nk}}. \tag{32}$$

En virtud de la suposición hecha sobre el rango de la matriz A_{i+1} las columnas a_k para $1 \leqslant k \leqslant M_2$ son linealmente independientes y el proceso de ortonormalización transcurre sin singularidades.

De (30)-(32) se deduce que las matrices A_{i+1} y B_{i+1} están conectadas por la relación $A_{i+1} = B_{i+1}\Omega_{i+1}$, donde Ω_{i+1} es la matriz triangular superior cuadrada de tamaño $M_2 \times M_2$ con elementos ω_{nk} para $1 \leqslant n \leqslant M_2$, $n \leqslant \leqslant k \leqslant M_2$ definidos en (30) y (32), y $\omega_{nk} = 0$ para $k < n$.

De esta forma, las fórmulas (30)-(32) determinan las matrices B_{i+1} y Ω_{i+1}. El cálculo no complicado muestra que

la construcción de las matrices B_{i+1} y Ω_{i+1} se puede realizar gastando: $MM_2^2 + 0{,}5\,(M_2^2 - M_2)$ operaciones de multiplicación, $MM_2^2 - M_2$ operaciones de suma y resta, MM_2 operaciones de división y M_2 extracciones de raíz cuadrada. Todo el proceso de ortonormalización indicado hay que realizarlo N veces en el paso directo del método de factorización. Esto exige $O\,(MNM_2^2)$ operaciones aritméticas y NM_2 operaciones de extracción de raíz cuadrada.

Nos queda indicar como se encuentran la matriz B_0 y el vector Y_0. Supondremos que las matrices P_{i+1} y Q_i para $0 \leqslant i \leqslant N-1$ no son degeneradas. Además, supongamos que la matriz P_0^{11} no es degenerada, y que la matriz Q_N tiene rango M_2.

Construyamos B_0 y Y_0. Sean $A_0 = \left\| \dfrac{(P_0^{11})^{-1} P_0^{12}}{E^{22}} \right\|$, $X_0 = \begin{pmatrix} (P_0^{11})^{-1} F_0 \\ 0 \end{pmatrix}$ la matriz triangular de tamaño $M \times M_2$ y el vector de dimensión M. Puesto que la dimensión de la matriz cuadrada unidad E^{22} es $M_2 \times M_2$, entonces el rango de A_0 es igual a M_2. La matriz B_0 se construye partiendo de A_0, con ayuda de ortonormalización (30)-(32), y Y_0 se elige por la fórmula $Y_0 = X_0 - B_0\varphi_0$ de la condición de ortogonalidad a las columnas de la matriz B_0, lo cual da $\varphi_0 = B_0^* X_0$. Ya que

$$B_0 = A_0\Omega_0^{-1}, \quad P_0 A_0 = \| P_0^{11} \,|\, - P_0^{12} \| \left\| \frac{(P_0^{11})^{-1} P_0^{12}}{E^{22}} \right\| = \| 0^{12} \|,$$

entonces $P_0 B_0 = 0^{12}$, de lo que $P_0 Y_0 = P_0 X_0 - P_0 B_0 \varphi_0 = P_0 X_0 = F_0$.

De esta forma, los B_0 y Y_0 construidos satisfacen las relaciones exigidas: $P_0 B_0 = 0^{12}$ y $P_0 Y_0 = F_0$.

Observemos que en virtud de la no degeneración de P_{i+1} y Q_i el rango de la matriz A_{i+1} coincide con el rango de B_i. Además, debido a la no degeneración de Ω_0 el rango de B_0 coincide con el rango de A_0 y es igual a M_2. Por lo tanto el proceso de ortonormalización (30)-(32) transcurre sin complicaciones. Luego, ya que los rangos de las matrices Q_N y B_N son iguales a M_2, entonces la matriz cuadrada $\Omega_{N+1} = Q_N B_N$ será no degenerada, lo que permite hallar el vector β_N.

De esta manera, el algoritmo del método de factorización ortogonal tiene la siguiente forma:

1) $B_i \Omega_i = A_i, \ i = 0, 1, 2, \ldots, N,$

$$A_i = P_i^{-1} Q_{i-1} B_{i-1}, \quad 1 \leqslant i \leqslant N, \quad A_0 = \left\| \frac{(P_0^{11})^{-1} P_0^{12}}{E^{22}} \right\|. \qquad (33)$$

Las matrices B_i y Ω_i para $0 \leqslant i \leqslant N$ se calculan por las fórmulas (30)-(32) y se guardan en la memoria. Se pone $\Omega_{N+1} = Q_N B_N$.

2) $Y_i = X_i - B_i \varphi_i$, $\varphi_i = B_i^* X_i$, $i = 0, 1, \ldots, N$, $X_i = P_i^{-1}(F_i + Q_{i-1} Y_{i-1})$, $1 \leqslant i \leqslant N$,

$$X_0 = \begin{pmatrix} (P_0^{11})^{-1} F_0 \\ 0 \end{pmatrix}. \tag{34}$$

Se calculan y se guardan en la memoria los vectores Y_i para $0 \leqslant i \leqslant N$ y φ_i para $1 \leqslant i \leqslant N$. Se pone $\varphi_{N+1} = Q_N Y_N$.

3) $\Omega_{i+1} \beta_i = \beta_{i+1} - \varphi_{i+1}$, $i = N, N-1, \ldots, 0$, $\beta_{N+1} = F_{N+1}$, $V_i = B_i \beta_i + Y_i$, $0 \leqslant i \leqslant N$. (35)

OBSERVACIÓN. Ya que las matrices Ω_i para $1 \leqslant i \leqslant N$ son las matrices triangulares superiores de tamaño $M_2 \times M_2$, entonces para encontrar β_i por β_{i+1} y φ_{i+1} dados, se exigen $O(M_2^2)$ operaciones.

Para ilustrar el algoritmo propuesto examinemos un ejemplo. Supongamos que se exige resolver el siguiente problema en diferencias tripuntual:

$$\begin{aligned} &-y_{i-1} + y_i - y_{i+1} = 0, \quad 1 \leqslant i \leqslant N-1, \\ &y_0 = 1, \quad y_N = 0. \end{aligned} \tag{36}$$

Este problema fue ya examinado por nosotros anteriormente en el punto 4 del § 2, donde por el método de factorización no monótona tripuntual fue hallada su solución para N no múltiplo de 3, y precisamente,

$$y_i = \frac{\operatorname{sen} \frac{(N-i)\pi}{3}}{\operatorname{sen} \frac{N\pi}{3}}, \quad 0 \leqslant i \leqslant N.$$

Reduzcamos el sistema (36) a un sistema de ecuaciones vectoriales bipuntuales del tipo (11), poniendo

$$V_i = \begin{pmatrix} y_i \\ y_{i+1} \end{pmatrix}, \quad 0 \leqslant i \leqslant N-1.$$

No es complejo ver, que (36) es equivalente al siguiente sistema:

$$\begin{aligned} &V_{i+1} - Q V_i = 0, \quad 0 \leqslant i \leqslant N-2, \\ &P_0 V_0 = 1, \quad Q_{N-1} V_{N-1} = 0, \end{aligned} \tag{37}$$

donde $P_0 = \| 1 | 0 \|$, $Q_{N-1} = \| 0 | 1 \|$, $Q = \left\| \frac{0 | 1}{-1 | 1} \right\|$. El sistema (37) es un caso particular de (11) con $M_1 = M_2 = 1$ y $M = 2$.

Para resolver (37) utilicemos el algoritmo de factorización ortogonal (33)-(35). Para el ejemplo examinado las matrices B_i tienen dimensión 2×1, Ω_i tiene dimensión 1×1, los vectores Y_i tendrán dimensión 2, y los vectores β_i y φ_i tienen dimensión 1.

En la tabla 3 se dan las matrices B_i y Ω_i y también los vectores Y_i, φ_i y β_i para $N = 11$. El método aplicado de factorización ortogonal permite obtener la solución exacta y_i del problema (36).

5. **Factorización para ecuaciones tripuntuales con coeficientes constantes.** Volvamos de nuevo al método de factorización matricial para ecuaciones tripuntuales y examinemos un caso particular de tales ecuaciones, precisamente:

$$\begin{gathered} -Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1, \\ Y_0 = F_0, \quad Y_N = F_N, \end{gathered} \tag{38}$$

donde C es la matriz cuadrada de tamaño $M \times M$, mientras Y_j y F_j son el vector buscado y el vector dado, ambos de dimensión M.

En el punto 1 fue mostrado, que a un sistema de ecuaciones tripuntuales del tipo (38) se reduce el problema de diferencias de Dirichlet para la ecuación de Poisson sobre una red rectangular, dada en un rectángulo, donde además la matriz C será simétrica y tridiagonal. A continuación, en el punto 2 del § 4 fue mostrado que el método de factorización matricial, el cual posee para (38) la forma

$$\alpha_{j+1} = (C - \alpha_j)^{-1}, \quad j = 1, 2, \ldots, N-1, \quad \alpha_1 = 0, \tag{39}$$

$$\beta_{j+1} = \alpha_{j+1}(F_j + \beta_j), \quad j = 1, 2, \ldots, N-1, \quad \beta_1 = F_0, \tag{40}$$

$$Y_j = \alpha_{j+1} Y_{j+1} + \beta_{j+1}, \quad j = N-1, N-2, \ldots, 1, \quad Y_N = F_N, \tag{41}$$

es correcto y estable. Allí fue mostrado que los valores propios de la matriz C son mayores que 2:

$$\lambda_k = \lambda_k(C) = 2 + 4\frac{h_2^2}{h_1^2}\operatorname{sen}^2\frac{k\pi h_1}{2l_1} > 2. \tag{42}$$

Recordemos, que en el caso de las ecuaciones vectoriales tripuntuales generales para el algoritmo de factorización

Tabla 3

i	0	1	2	3	4	5	6	7	8	9	10	11
Ω_i	1	$\sqrt{2}$	$\frac{1}{\sqrt{2}}$	1	$\sqrt{2}$	$\frac{1}{\sqrt{2}}$	1	$\sqrt{2}$	$\frac{1}{\sqrt{2}}$	1	$\sqrt{2}$	$-\frac{1}{\sqrt{2}}$
φ_i	0	$-\frac{1}{\sqrt{2}}$	$-\frac{1}{2}$	1	$-\frac{1}{\sqrt{2}}$	$-\frac{1}{2}$	1	$-\frac{1}{\sqrt{2}}$	$-\frac{1}{2}$	1	$-\frac{1}{\sqrt{2}}$	$\frac{1}{2}$
β_i	1	$\frac{1}{\sqrt{2}}$	0	1	$\frac{1}{\sqrt{2}}$	0	1	$\frac{1}{\sqrt{2}}$	0	1	$\frac{1}{\sqrt{2}}$	0
B_i	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$	$\begin{pmatrix} -1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$	
Y_i	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} -1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} -1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$	
y_i	1	1	0	−1	−1	0	1	1	0	−1	−1	0

matricial se exigen $O(M^3N)$ operaciones aritméticas para calcular las matrices α_j y $O(M^2N)$ operaciones para el cálculo de los vectores de factorización β_j y de la solución Y_j. Para conservar las matrices completas α_j, en general no simétricas, es necesario guardar en la memoria $M^2(N+1)$ elementos de estas matrices. ¿Disminuyen o no estas cantidades, si con el método de factorización matricial se resuelve un sistema vectorial tripuntual (38) especial con coeficientes constantes?

Para el ejemplo examinado todas las matrices α_j serán simétricas en virtud de la simetría de la matriz C, pero aunque C es la matriz tridiagonal, todas las matrices α_j, $j \geqslant 2$, serán completas. Por consiguiente, teniendo en cuenta la simetría de las matrices α_j, se puede disminuir solamente el volumen de información intermedia recordada, pero no en más de dos veces. El orden en M y N del número de operaciones aritméticas no cambia.

Construyamos ahora una modificación del algoritmo (39)-(41), la cual no exige memoria complementaria para conservar la información intermedia y se realiza con un gasto de $O(MN^2)$ operaciones aritméticas, si se resuelve el problema (38) con una matriz tridiagonal C.

Primeramente hallemos la forma explícita de las matrices de factorización α_j para cualquier j. Para eso, aplicando (39), expresemos α_j mediante la matriz C.

Observando, que

$$\alpha_1 = 0, \quad \alpha_2 = C^{-1}, \quad \alpha_3 = (C^2 - E)^{-1}C, \tag{43}$$

buscaremos la solución de la ecuación no lineal de diferencias (39) en la forma

$$\alpha_j = P_{j-1}^{-1}(C)\, P_{j-2}(C), \quad j \geqslant 2,$$

donde $P_j(C)$ es el polinomio de C de grado j. Escribamos de nuevo a (39) en la forma

$$\alpha_{j+1}(C - \alpha_j) = E, \quad j \geqslant 2,$$

y sustituyamos aquí (44). Así obtendremos la relación recurrente $P_j(C) = CP_{j-1}(C) - P_{j-2}(C)$, $j \geqslant 2$, o después del desplazamiento del índice en una unidad y teniendo en cuenta (43)

$$P_{j+1}(C) = CP_j(C) - P_{j-1}(C), \quad j \geqslant 1, \qquad P_0(C) = E, \quad P_1(C) = C. \tag{45}$$

De esta manera, las fórmulas (45) definen completamente el polinomio $P_j(C)$ para cualquier $j \geqslant 0$.

Hallemos la solución de (45). El polinomio algebraico correspondiente satisface las relaciones

$$P_{j+1}(t) = tP_j(t) - P_{j-1}(t), \quad j \geqslant 1,$$
$$P_0(t) = 1, \quad P_1(t) = t,$$

las cuales representan el problema de Cauchy para una ecuación de diferencias tripuntual con coeficientes constantes. En el punto 2 del § 4, del cap. I fue hallada la solución de este problema $P_j(t) = U_j\left(\frac{t}{2}\right)$, $j \geqslant 0$, donde $U_j(x)$ es el polinomio de Chebishev de segundo género y de grado j

$$U_j(x) = \begin{cases} \dfrac{\operatorname{sen}((j+1)\arccos x)}{\operatorname{sen}\arccos x}, & |x| \leqslant 1, \\ \dfrac{\operatorname{sh}((j+1)\operatorname{Arch} x)}{\operatorname{sh}\operatorname{Arch} x}, & |x| \geqslant 1. \end{cases}$$

De este modo, está hallada la expresión explícita para las matrices de factorización α_j:

$$\alpha_j = U_{j-1}^{-1}\left(\frac{C}{2}\right) U_{j-2}\left(\frac{C}{2}\right), \quad j \geqslant 2, \quad \alpha_1 = 0. \tag{46}$$

Esto nos libra de la necesidad de realizar los cálculos de las matrices de factorización α_j según la fórmula (39), para lo que se exige el volumen fundamental de trabajo computacional en el algoritmo (39)-(41). Además no hay necesidad de recordar las matrices α_j.

Examinemos ahora las fórmulas (40) y (41). Ellas contienen el producto de la matriz α_{j+1} por los vectores $F_j + \beta_j$ y Y_{j+1}. Mostremos ahora como se puede determinar el producto de la matriz α_j por un vector, sin calcular α_j por la fórmula (46). Para eso necesitaremos el lema 6, el cual citaremos sin demostración.

LEMA 6. *Supongamos que el polinomio $f_n(x)$ de grado n posee raíces simples. El cociente entre el polinomio $g_m(x)$ de grado m y el polinomio $f_n(x)$ de grado $n > m$ sin raíces comunes puede ser representado en forma de la suma de n fracciones elementales*

$$\frac{g_m(x)}{f_n(x)} = \sum_{l=1}^{n} \frac{a_l}{x - x_l}, \quad a_l = \frac{g_m(x_l)}{f'_n(x_l)},$$

donde x_l son las raíces de $f_n(x)$, y $f'_n(x)$ es la derivada del polinomio $f_n(x)$.

Utilizando el lema 6, hallaremos el desarrollo en fracciones simples del cociente $\varphi(x) = \frac{U_{j-2}(x)}{U_{j-1}(x)}$, $j \geqslant 2$. Como las raíces de $U_{j-1}(x)$ son

$$x_k = \cos\frac{k\pi}{j}, \quad k = 1, \quad 2, \quad \ldots, \quad j-1,$$

y

$$U_{j-2}(x_k) = (-1)^{k-1}, \quad \frac{d}{dx}[U_{j-1}(x_k)] = \frac{j(-1)^{k-1}}{\operatorname{sen}^2 \frac{k\pi}{j}},$$

entonces en virtud del lema 6 tendremos el siguiente desarrollo para $\varphi(x)$:

$$\varphi(x) = \frac{U_{j-2}(x)}{U_{j-1}(x)} = \sum_{k=1}^{j-1} \frac{\operatorname{sen}^2 \frac{k\pi}{j}}{j} \left(x - \cos\frac{k\pi}{j}\right)^{-1}. \tag{47}$$

De (46) y (47) se deduce una representación más para las matrices α_j la cual nosotros utilizaremos

$$\alpha_j = \sum_{k=1}^{j-1} a_{kj} \left(C - 2\cos\frac{k\pi}{j} E\right)^{-1},$$

$$a_{kj} = \frac{2\operatorname{sen}^2 \frac{k\pi}{j}}{j}, \quad j \geqslant 2. \tag{48}$$

Usando (48) se puede realizar el producto de la matriz α_j por el vector Y según el siguiente algoritmo: para $k = 1, 2, \ldots, j-1$ se resuelven las ecuaciones

$$\left(C - 2\cos\frac{k\pi}{j} E\right) V_k = a_{kj} Y, \tag{49}$$

donde a_{kj} está definido en (48) el resultado $\alpha_j Y$ se obtiene por sumación sucesiva de los vectores V_k

$$\alpha_j Y = \sum_{k=1}^{j-1} V_k. \tag{50}$$

Notemos, que en virtud de (42) la matriz $C - 2\cos\frac{k\pi}{j} E$ es no degenerada y, además, tridiagonal, si así lo era la matriz C. En este caso cada una de las ecuaciones (49) se resuelve en $O(M)$ operaciones aritméticas por el método de factorización tripuntual, descrito en el § 1. Por consiguiente, en la resolución de todos los problemas (49) y además

en el cálculo de la suma (50) se exigen $O(Mj)$ operaciones. Puesto que en (40) y (41) el producto de la matriz α_j por vectores se efectúa para $j = 2, 3, \ldots, N$, entonces el método modificado de factorización matricial (40), (41), y (49), (50) exige $O(MN^2)$ operaciones aritméticas.

De esta forma, hemos construido el método modificado de factorización matricial, que permite hallar la solución del problema de Dirichlet de diferencias para la ecuación de Poisson en un rectángulo con un gasto de $O(MN^2)$ operaciones aritméticas. La disminución del número de operaciones en comparación con el algoritmo inicial (39)-(41) se logra por cuenta del carácter específico del problema a resolver.

En los dos capítulos siguientes nosotros examinaremos otros métodos directos de resolución del problema indicado y de otros problemas de diferencias semejantes, los cuales exigirán un número menor aún de operaciones que el método aquí construido.

Capítulo III
Método de reducción completa

En este capítulo se estudia un método de resolución de ecuaciones elípticas reticulares especiales: el método de reducción completa. Este método directo permite hallar la solución del problema de Dirichlet de diferencias para la ecuación de Poisson en un rectángulo en $O(N^2 \log_2 N)$ operaciones aritméticas, donde N es el número de nodos de la red por cada dirección.

En el § 1 está dado el planteamiento de los problemas de contorno para las ecuaciones en diferencias, en cuya resolución se puede utilizar el método de reducción. En el § 2 se expone el algoritmo del método para el caso del primer problema de contorno y en el § 3 son examinados ejemplos de aplicación del método. En el § 4 está dada una generalización del método al caso de condiciones de contorno generales.

§ 1. Problemas de contorno para ecuaciones vectoriales tripuntuales

1. Planteamiento de los problemas de contorno. En el capítulo II fueron construidos los métodos de factorización escalar y matricial para resolver ecuaciones tripuntuales escalares y vectoriales. El método de factorización matricial para una ecuación con coeficientes variables se realiza con un gasto de $O(M^3N)$ operaciones aritméticas, donde N es el número de ecuaciones, y M es la dimensión de los vectores de las incógnitas (el número de incógnitas en el problema es igual a MN). Para ciertas clases especiales de ecuaciones vectoriales, correspondientes, por ejemplo, al problema de Dirichlet de diferencias para la ecuación de Poisson en un rectángulo, fue propuesto el algoritmo modificado del método de factorización matricial. Este algoritmo permite reducir el número de operaciones hasta $O(MN^2)$.

Este capítulo está dedicado al ulterior estudio de los métodos directos de solución de ecuaciones vectoriales espe-

ciales, a las cuales se reducen los esquemas de diferencias para las ecuaciones elípticas más simples. Será construido el método de reducción completa que permite resolver los problemas de contorno fundamentales con un gasto de $O(MN \log_2 N)$ operaciones aritméticas. Si no tenemos en cuenta la débil dependencia logarítmica de N, entonces el número de operaciones para este método es proporcional al número de incógnitas MN. La creación de este método es un paso esencial en el desarrollo tanto de los métodos directos como de los métodos iterativos de solución de ecuaciones reticulares.

Formulemos los problemas de contorno para ecuaciones vectoriales tripuntuales cuya solución se puede hallar por el método de reducción completa. Nosotros examinaremos los siguientes problemas:

1) *Primer problema de contorno.* Se exige hallar la solución de la ecuación

$$-Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1, \quad (1)$$

que satisfaga valores dados para $j = 0$ y $j = N$.

$$Y_0 = F_0, \quad Y_N = F_N. \quad (2)$$

Aquí Y_j es el vector de las incógnitas de número j, F_j es un miembro derecho dado y C una matriz cuadrada dada.

2) *Segundo y tercer problemas de contorno.* Se busca la solución de la ecuación (1), que satisfaga las siguientes condiciones de contorno para $j = 0$ y $j = N$:

$$(C + 2\alpha E)\, Y_0 - 2Y_1 = F_0, \quad j = 0,$$
$$-2Y_{N-1} + (C + 2\beta E)\, Y_N = F_N, \quad j = N, \quad (3)$$

donde $\alpha \geqslant 0$ y $\beta \geqslant 0$. Para $\alpha = \beta = 0$ las fórmulas (3) prefijan las condiciones de contorno de segundo género. Nosotros también examinaremos combinaciones de condiciones de contorno, por ejemplo, cuando para $j = 0$ está dada una condición de contorno de primer género y para $j = N$ de tercero o de segundo género.

3) *Problema de contorno periódico.* Se exige hallar la solución de la ecuación $-Y_{j-1} + CY_j - Y_{j+1} = F_j$, que es periódica, o sea, $Y_{N+j} = Y_j$. Se supone que el miembro derecho F_j también es periódico, $F_{N+j} = F_j$. Este problema se formula en la siguiente forma equivalente: hallar la solución de la ecuación

$$-Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1,$$
$$-Y_{N-1} + CY_0 - Y_1 = F_0, \quad Y_N = Y_0. \quad (4)$$

A las ecuaciones de tal género se reducen los esquemas de diferencias para ecuaciones elípticas en sistemas de coordenadas curvilíneas ortogonales: en sistemas cilíndricos, polares y esféricos.

Además de la ecuación vectorial fundamental (1), que contiene una matriz C, nosotros examinaremos el primer problema de contorno para la ecuación más general

$$-BY_{j-1} + AY_j - BY_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1,$$
$$Y_0 = F_0, \quad Y_N = F_N \tag{5}$$

con matrices cuadradas A y B. Semejante tipo de problemas aparece al resolver el problema de Dirichlet de diferencias de un orden de exactitud aumentado para la ecuación de Poisson en un rectángulo.

Formulemos las exigencias sobre las matrices C, A y B, que aseguran la posibilidad de aplicación del método de reducción completa para la resolución de los problemas planteados (1)-(5). Para los problemas (1)-(4) supondremos, que para cualquier vector Y es válida la desigualdad $(CY, Y) \geqslant 2\,(Y, Y)$, y para el problema (5) la desigualdad $(AY, Y) \geqslant 2\,(BY, Y) > 0$. Aquí se utiliza el producto escalar de vectores usual.

2. Primer problema de contorno. Comenzaremos el estudio del método de reducción completa por la descripción de los problemas de contorno reticulares para ecuaciones elípticas, que pueden ser escritos en la forma de las ecuaciones vectoriales especiales (1)-(5). Supongamos que sobre la red rectangular

$$\bar{\omega} = \{x_{ij} = (ih_1, jh_2) \in \bar{G}, \quad 0 \leqslant i \leqslant M,$$
$$0 \leqslant j \leqslant N, \quad h_1 = l_1/M, \quad h_2 = l_2/N\}$$

con frontera γ introducida en el rectángulo $\bar{G} = \{0 \leqslant x_\alpha \leqslant$ $\leqslant l_\alpha$, $\alpha = 1, 2\}$, se exige hallar la solución del problema de Dirichlet de diferencias para la ecuación de Poisson

$$\begin{aligned} &y_{\bar{x}_1 x_1} + y_{\bar{x}_2 x_2} = -\varphi(x), \quad x \in \omega \\ &y(x) = g(x), \qquad x \in \gamma. \end{aligned} \tag{6}$$

En el § 4 del cap. II fue mostrado que el problema (6) puede ser escrito en la forma (1), (2), donde Y_j es el vector de dimensión $M-1$ cuyas componentes son los valores de la función reticular $y(i, j) = y(x_{ij})$ en los nodos inte-

riores de la j-ésima fila de la red $\bar{\omega}$:

$$Y_j = (y(1, j)\ y(2, j), \ldots, y(M-1, j)), \quad 0 \leqslant j \leqslant N.$$

C es la matriz cuadrada de dimensión $(M-1) \times (M-1)$, la cual corresponde al operador de diferencias Λ, donde

$$\begin{aligned} &\Lambda y = 2y - h_2^2 y_{\bar{x}_1 x_1}, \quad h_1 \leqslant x_1 \leqslant l_1 - h_1, \\ &y = 0, \quad x_1 = 0, \ l_1. \end{aligned} \tag{7}$$

El miembro derecho F_j es el vector de dimensión $M-1$ definido de la siguiente manera:

1) para $j = 1, 2, \ldots, N-1$

$$F_j = (h_2^2 \bar{\varphi}(1, j),\ h_2^2 \varphi(2, j), \ldots \\ \ldots, h_2^2 \varphi(M-2, j),\ h_2^2 \bar{\varphi}(M-1, j)), \tag{8}$$

donde

$$\bar{\varphi}(1, j) = \varphi(1, j) + \frac{1}{h_1^2} g(0, j),$$

$$\bar{\varphi}(M-1, j) = \varphi(M-1, j) + \frac{1}{h_1^2} g(M, j);$$

2) para $j = 0, N$

$$F_j = (g(1, j),\ g(2, j), \ldots, g(M-1, j)). \tag{9}$$

De (7) se deduce que para el ejemplo examinado la matriz C es la matriz tridiagonal simétrica.

Examinemos un problema de diferencias más complejo, que también se escribe en la forma de las ecuaciones (1), (2). Supongamos que sobre la red $\bar{\omega}$ se exige hallar la solución de la ecuación de Poisson en diferencias

$$y_{\bar{x}_1 x_1} + y_{\bar{x}_2 x_2} = -\varphi(x), \quad x \in \omega, \tag{10}$$

que satisface en los lados $x_1 = 0$ y $x_1 = l_1$ las condiciones de contorno de tercero o de segundo género,

$$\frac{2}{h_1} y_{x_1} + y_{\bar{x}_2 x_2} = \frac{2}{h_1} \varkappa_{-1} y - \bar{\varphi}, \quad x_1 = 0, \tag{11}$$

$$-\frac{1}{h_1} y_{\bar{x}_1} + y_{\bar{x}_2 x_2} = \frac{2}{h_1} \varkappa_{+1} y - \bar{\varphi}, \quad x_1 = l_1, \tag{12}$$

$$h_2 \leqslant x_2 \leqslant l_2 - h_2$$

y las condiciones de contorno de primer género en los lados $x_2 = 0$ y $x_2 = l_2$: $y(x) = g(x)$, $x_2 = 0$, l_2, $0 \leqslant x_1 \leqslant l_1$. Para que el problema planteado pueda ser escrito en la

forma (1), (2) con la matriz C no dependiente de j, es necesario suponer que se cumple la condición $\varkappa_{\pm 1}$ = const.

Reduzcamos este problema a (1), (2). Para eso multipliquemos (10)-(12) por $(-h_2^2)$ y anotemos la derivada de diferencias $y_{\bar{x}_2 x_2}$ por puntos para todo $j = 1, 2, \ldots, N-1$. Obtendremos las siguientes ecuaciones:

1) para $i = 0$

$$-y(0, j-1) + 2\left[\left(1 + \frac{h_2^2}{h_1}\varkappa_{-1}\right) y(0, j) - \right.$$
$$\left. - \frac{h_2^2}{h_1} y_{x_1}(0, j)\right] - y(0, j+1) = h_2^2\bar{\varphi}(0, j);$$

2) para $i = 1, 2, \ldots, M-1$

$$-y(i, j-1) + [2y(i, j) - h_2^2 y_{\bar{x}_1 x_1}(i, j)] -$$
$$- y(i, j+1) = h_2^2\varphi(i, j);$$

3) para $i = M$

$$-y(M, j-1) + 2\left[\left(1 + \frac{h_2^2}{h_1}\varkappa_{+1}\right) y(M, j) + \frac{h_2^2}{h_1} y_{\bar{x}_1}(M, j)\right] -$$
$$- y(j, j+1) = h_2^2\varphi(i, j)$$

Designemos

$$Y_j = (y(0, j), y(1, j), \ldots, y(M, j)), \quad 0 \leqslant j \leqslant N,$$
$$F_j = (h_2^2\bar{\varphi}(0, j), h_2\varphi(1, j), \ldots, h_2^2\bar{\varphi}(M-1, j),$$
$$h_2^2\bar{\varphi}(M, j)), \qquad (13)$$
$$1 \leqslant j \leqslant N-1$$
$$F_j = (g(0, j), g(1, j), \ldots, g(M, j)), \quad j = 0, N.$$

En estas notaciones las ecuaciones obtenidas se escriben en la forma (1), (2) donde la matriz cuadrada C de dimensión $(M+1) \times (M+1)$ corresponde al operador de diferencias Λ:

$$\Lambda y = \begin{cases} 2\left(1 + \frac{h_2^2}{h_1}\varkappa_{-1}\right) y - \frac{2h_2^2}{h_1} y_{x_1}, & x_1 = 0, \\ 2y - h_2^2 y_{\bar{x}_1 x_1}, & h_1 \leqslant x_1 \leqslant l_1 - h_1, \\ 2\left(1 + \frac{h_2^2}{h_1}\varkappa_{+1}\right) y + 2\frac{h_2^2}{k_1} y_{\bar{x}_1}, & x_1 = l_1. \end{cases} \qquad (14)$$

Aquí de nuevo tenemos que ver con el caso, cuando C es una matriz tridiagonal. El dar sobre los lados $x_1 = 0, l_1$ condiciones de contorno de tercer género (11), (12) en lugar de condiciones de primer género conduce solamente a otra

definición del operador Λ: en lugar de (7) tenemos (14). A su vez el tipo de la ecuación (1) y de las condiciones de contorno no cambia. Si para $x_1 = 0$ en lugar de la condición (11) es dada la condición de contorno de primer género $y(x) = g(x)$ y para $x_1 = l_1$ como antes se da la condición (12), entonces este problema en diferencias también se reduce a (1), (2). En este caso

$Y_j = (y(1, j), y(2, j), \ldots, y(M, j)), \quad 0 \leqslant j \leqslant N,$

$F_j = (h_2^2\bar{\varphi}(1, j), h_2^2\varphi(2, j), \ldots, h_2^2\varphi(M-1, j), h_2^2\bar{\varphi}(M, j)) \quad 1 \leqslant j \leqslant N-1,$

donde $\bar{\varphi}(1, j) = \varphi(1, j) + \frac{1}{h_1^2} g(0, j)$, $\bar{\varphi}(M, j)$ es el valor del miembro derecho $\bar{\varphi}$ de (12) en el punto correspondiente, y la matriz cuadrada C corresponde al operador de diferencias Λ, siendo

$$\Lambda y = \begin{cases} 2y - h_2^2 y_{\bar{x}_1 x_1}, & h_1 \leqslant x_1 \leqslant l_1 - h_1, \\ 2\left(1 + \frac{h_2^2}{h_1}\varkappa_{+1}\right) y + 2\frac{h_2^2}{h_1} y_{\bar{x}_1}, & x_1 = l_1 \end{cases} \qquad (15)$$

y $y = 0$ para $x_1 = 0$.

Si está dada la condición de contorno de primer género para $x_1 = l_1$, y la condición de contorno de tercer género (11) para $x_1 = 0$, entonces en (1), (2)

$Y_j = (y(0, j), y(1, j), \ldots, y(M-1, j)), \quad 0 \leqslant j \leqslant N,$

$F_j = (h_2^2\bar{\varphi}(0, j), h_2^2\varphi(1, j), \ldots, h_2^2\varphi(M-2, j), h_2^2\bar{\varphi}(M-1, j)),$

$1 \leqslant j \leqslant N-1,$

donde $\bar{\varphi}(M-1, j) = \varphi(M-1, j) + \frac{1}{h_1^2} g(M, j)$, y la matriz C corresponde al operador de diferencias Λ, donde

$$\Lambda y = \begin{cases} 2\left(1 + \frac{h_2^2}{h_1}\varkappa_{-1}\right) y - \frac{2h_2^2}{h_1} y_{x_1}, & x_1 = 0, \\ 2y - h_2^2 y_{\bar{x}_1 x_1}, & h_1 \leqslant x_1 \leqslant l_1 - h_1 \end{cases} \qquad (16)$$

y $y = 0$ para $x_1 = l_1$.

De esta forma hemos mostrado, que si por la dirección x_2 están dadas condiciones de contorno de primer género y por la dirección x_1 combinaciones cualesquiera de condiciones de contorno de primero, segundo o tercer género, entonces los esquemas de diferencias para la ecuación de

Poisson en un rectángulo se escriben en la forma del primer problema de contorno para ecuaciones vectoriales tripuntuales (1), (2). La matriz C se determina con ayuda del operador de diferencias Λ, el cual en dependencia del tipo de la condición de contorno sobre los lados $x_1 = 0$ y $x_1 = l_1$ se da por las fórmulas (7), (14)-(16).

3. Otros problemas de contorno para ecuaciones en diferencias. El tipo de las condiciones de contorno para la ecuación (1) se determina completamente por el tipo de las condiciones de contorno para la ecuación en diferencias (10) sobre los lados del rectángulo $x_2 = 0$ y $x_2 = l_2$. Nosotros examinamos el caso, cuando sobre estos lados fueron dadas condiciones de contorno de primer género.

Examinemos ahora otros problemas de contorno para la ecuación (10), los cuales se reducen a las ecuaciones vectoriales (1), (3). Supongamos que sobre la red rectangular $\bar{\omega}$ definida más arriba, se exige hallar la solución *del tercer problema de contorno* para la ecuación de Poisson en diferencias. El esquema de diferencias tiene la forma siguiente:

$$y_{\bar{x}_1 x_1} + y_{\bar{x}_2 x_2} = -\varphi(x), \quad x \in \omega, \tag{17}$$

$$\frac{2}{h_1} y_{x_1} + y_{\bar{x}_2 x_2} = \frac{2}{h_1} \varkappa_{-1} y - \bar{\varphi}, \quad x_1 = 0, \tag{18}$$

$$-\frac{2}{h_1} y_{\bar{x}_1} + y_{\bar{x}_2 x_2} = \frac{2}{h_1} \varkappa_{+1} y - \bar{\varphi}, \quad x_1 = l_1,$$

$$h_2 \leqslant x_2 \leqslant l_2 - h_2,$$

$$y_{\bar{x}_1 x_1} + \frac{2}{h^2} y_{x_2} = \frac{2}{h_2} \varkappa_{-2} y - \bar{\varphi}, \quad x_2 = 0, \tag{19}$$

$$y_{\bar{x}_1 x_1} - \frac{2}{h_2} y_{\bar{x}_2} = \frac{2}{h_2} \varkappa_{+2} y - \bar{\varphi}, \quad x_2 = l_2,$$

$$h_1 \leqslant x_1 \leqslant l_1 - h_1. \tag{20}$$

La aproximación en las esquinas de la red posee la forma especial:

$$\frac{2}{h_1} y_{x_1} + \frac{2}{h_2} y_{x_2} = \left(\frac{2}{h_1} \varkappa_{-1} + \frac{2}{h_2} \varkappa_{-2} \right) y - \bar{\varphi}, \quad x_1 = 0, \quad x_2 = 0, \tag{21}$$

$$-\frac{2}{h_1} y_{\bar{x}_1} + \frac{2}{h_2} y_{x_2} = \left(\frac{2}{h_1} \varkappa_{+1} + \frac{2}{h_2} \varkappa_{-2} \right) y - \bar{\varphi}, \quad x_1 = l_1, \quad x_2 = 0, \tag{22}$$

$$\frac{2}{h_1} y_{x_1} - \frac{2}{h_2} y_{\bar{x}_2} = \left(\frac{2}{h_1} \varkappa_{-1} + \frac{2}{h_2} \varkappa_{+2}\right) y - \bar{\varphi}, \quad x_1 = 0, \quad x_2 = l_2, \tag{23}$$

$$-\frac{2}{h_1} y_{\bar{x}_1} - \frac{2}{h_2} y_{\bar{x}_2} = \left(\frac{2}{h_1} \varkappa_{+1} + \frac{2}{h_2} \varkappa_{+2}\right) y - \bar{\varphi}, \quad x = l_1, \quad x_2 = l_2, \tag{24}$$

Aquí se presupone, que se cumplen las condiciones $\varkappa_{\pm\alpha} = \text{const}$, $\alpha = 1, 2$.

Mostremos que el problema (17)—(24) se reduce a (1), (3). En efecto, designando por Y_j el vector de dimensión $M + 1$,

$$Y_j = (y(0, j), y(1, j), \ldots, y(M, j)), \quad 0 \leqslant j \leqslant N$$

y definiendo el miembro derecho F_j para $j = 1, 2, \ldots, \ldots, N - 1$ según las fórmulas (13) obtendremos de (17) y (18), como en el punto anterior, las ecuaciones (1) con la matriz C correspondiente a Λ de (14). Queda por mostrar, que las condiciones (19)—(24) pueden ser escritas en forma de las condiciones de contorno (3).

Multipliquemos (19), (21) y (22) por $(-h_2^2)$ y anotemos por puntos la derivada de diferencias y_{x_2} que entra en ellas. Obtenemos:

1) para $i = 0$

$$2\left[\left(1 + \frac{h_2^2}{h_2} \varkappa_{-1}\right) y(0, 0) - \frac{h_2^2}{h_1} y_{x_1}(0, 0)\right] + \\ + 2h_2\varkappa_{-2} y(0, 0) - 2y(0, 1) = h_2^2\bar{\varphi}(0, 0),$$

2) para $i = 1, 2, \ldots, M - 1$

$$2y(i, 0) - h_2^2 y_{\bar{x}_1 x_1}(i, 0) + 2h_2\varkappa_{-2} y(i, 0) - 2y(i, 1) = h_2^2\bar{\varphi}(i, 0),$$

3) para $i = M$

$$2\left[\left(1 + \frac{h_2^2}{h_1} \varkappa_{+1}\right) y(M, 0) + \frac{h_2^2}{h_1} y_{\bar{x}_1}(M, 0)\right] + \\ + 2h_2\varkappa_{-2} y(M, 0) - 2y(M, 1) = h_2^2\bar{\varphi}(M, 0).$$

Si denotamos $\alpha = h_2\varkappa_{-2}$, entonces estas igualdades se pueden escribir en la forma vectorial

$$(C + 2\alpha E)(Y_0 - 2Y_1 = F_0, \tag{25}$$

donde $F_0 = (h_2^2\bar{\varphi}(0, 0), h_2^2\bar{\varphi}(1, 0), \ldots, h_2^2\bar{\varphi}(M, 0))$.

Análogamente de (20), (23) y (24) se encuentra la ecuación

$$-2Y_{N-1} + (C + 2\beta E) Y_N = F_N,$$

donde se denota $\beta = h_2\varkappa_{+2}$ y $F_N = (h_2^2\bar{\varphi}(0, N), h_2^2\bar{\varphi}(1, N), \ldots, h_2^2\bar{\varphi}(M, N))$. De esta manera, el esquema de diferencias (17)—(24) se reduce al problema (1), (3).

Examinemos ahora el caso de prefijar ciertas combinaciones de condiciones de contorno sobre los lados del rectángulo $\bar{G}$. Como fue señalado más arriba, al prefijar condiciones de contorno diferentes de (18) sobre los lados $x_1 = 0$ y $x_1 = l_1$, influye solamente en la definición de la matriz C. Si para $x_2 = 0$ está prefijada la condición de contorno de primer género, es decir, en lugar de (19), (21) y (22) se da $y(x) = g(x)$, $x_2 = 0$, entonces la condición (25) debe ser sustituida por la condición $Y_0 = F_0$, donde $F_0 = (g(0, 0), \ldots, g(M, 0))$. En este caso el problema de contorno vectorial tripuntual posee la forma

$$\begin{gathered} -Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1, \\ Y_0 = F_0, \\ -2Y_{N-1} + (C + 2\beta E)\, Y_N = F_N. \end{gathered} \tag{26}$$

A un sistema análogo llegamos también en el caso, cuando sobre el lado $x_2 = l_2$ está dada una condición de contorno de primer género y sobre el lado $x_2 = 0$ una condición de contorno de tercer género. En este caso el problema de contorno vectorial tiene la forma

$$\begin{gathered} -Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1, \\ (C + 2\alpha E)\, Y_0 - 2Y_1 = F_0, \quad Y_N = F_N. \end{gathered} \tag{27}$$

Hemos examinado ejemplos de problemas de contorno para la ecuación de Poisson en diferencias en un rectángulo y mostramos, que a ellos corresponden los problemas de contorno (1), (2) ó (1), (3), ó (26), (27) con su correspondiente matriz diagonal C.

A los problemas de contorno vectoriales indicados se reducen también los esquemas de diferencias para ecuaciones elípticas más complejas tanto en coordenadas cartesianas como en sistemas de coordenadas curvilíneos ortogonales. Citemos ejemplos. En un sistema cartesiano estos son los problemas de contorno fundamentales para la ecuación elíptica

$$\frac{\partial}{\partial x_1}\left(k_1(x_1)\frac{\partial u}{\partial x_1}\right) + k_2(x_1)\frac{\partial^2 u}{\partial x_2^2} - q(x_1)\, u = -f(x), \quad x \in G,$$

cuyos coeficientes dependen solamente de una variable. En este caso en el rectángulo $\bar{G}$ se puede introducir una red

rectangular $\bar{\omega}$ con paso uniforme h_2 en la dirección de x_2 y con pasos no uniformes arbitrarios en la dirección de x_1.

En un sistema de coordenadas cilíndricas tales ejemplos son los problemas de contorno para la ecuación de Poisson en un cilindro circular finito o en un tubo en presencia de simetría axial:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)+\frac{\partial^2 u}{\partial z^2}=-f(r, z),$$
$$0\leqslant r_0<r<R, \quad 0<z<l.$$

En este caso por la dirección de r se puede introducir una red no uniforme arbitraria y por la dirección de z una red con paso constante h_2.

Si para la ecuación de Poisson se plantea el problema de buscar la solución sobre la superficie del cilindro, es decir

$$\frac{1}{R^2}\frac{\partial^2 u}{\partial \varphi^2}+\frac{\partial^2 u}{\partial z^2}=-f(\varphi, z), \quad 0\leqslant\varphi\leqslant 2\pi, \quad 0<z<l,$$

entonces el problema de diferencias correspondiente se reduce al problema de contorno vectorial periódico (4), al mismo tiempo por la dirección z se admite una red arbitraria no uniforme.

En un sistema polar de coordenadas son admisibles los esquemas de diferencias para la ecuación de Poisson en un círculo, un anillo y en sectores circulares o anulares

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 u}{\partial \varphi^2}=-f(r, \varphi), \quad (r, \varphi)\in G.$$

Para el círculo y el anillo el esquema de diferencias se reduce al problema periódico (4), y para los sectores se reduce a los problemas (1), (2) ó (1), (3). Aquí se puede introducir una red no uniforme por la dirección de r.

Al problema de contorno periódico (4) se reduce el esquema de diferencias para la ecuación de Poisson, dada sobre la superficie de la esfera de radio R:

$$\frac{1}{R^2\operatorname{sen}\theta}\frac{\partial}{\partial\theta}\left(\operatorname{sen}\theta\,\frac{\partial u}{\partial\theta}+\frac{1}{R^2\operatorname{sen}^2\theta}\frac{\partial^2 u}{\partial\varphi^2}\right)=-f(\varphi, \theta).$$

4. **Problema de Dirichlet de diferencias con orden de exactitud aumentado.** Examinemos ahora un ejemplo de esquema de diferencias, el cual se reduce a la ecuación vectorial (5) más general que (1). Escribamos sobre la red rectangular $\bar{\omega}=\{x_{ij}=(ih_1, jh_2)\in\bar{G}, 0\leqslant i\leqslant M, 0\leqslant j\leqslant N, h_1M=l_1, h_2N=l_2\}$ el problema de Dirichlet de diferencias para

la ecuación de Poisson con orden de exactitud aumentado

$$y_{\bar{x}_1 x_1} + y_{\bar{x}_2 x_2} + \frac{h_1^2 + h_2^2}{12} y_{\bar{x}_1 x_1 \bar{x}_2 x_2} = -\varphi(x), \quad x \in \omega, \tag{28}$$
$$y(x) = g(x), \quad x \in \gamma.$$

Si $h_1 \neq h_2$, la solución del esquema de diferencias (28) para la correspondiente elección del segundo miembro $\varphi(x)$, converge con velocidad $O(h_1^4 + h_2^4)$ a una solución suficientemente suave del problema diferencial, y con velocidad $O(h^6)$, si $h_1 = h_2 = h$.

Reduzcamos (28) a un problema de contorno para la ecuación vectorial tripuntual

$$-BY_{j-1} + AY_j - BY_{j+1} = F_j, \qquad 1 \leqslant j \leqslant N-1,$$
$$Y_0 = F_0, \quad Y_N = F_N. \tag{29}$$

Para esto es necesario multiplicar (28) por $(-h_2^2)$, anotar la derivada de diferencias $\left(y + \frac{h_1^2 + h_2^2}{12} y_{\bar{x}_1 x_1}\right)_{\bar{x}_2 x_2}$ por puntos y utilizar las notaciones

$$Y_j = (y(1, j), y(2, j), \ldots, y(M-1, j)),$$
$$F_j = (h_2^2 \bar{\varphi}(1, j), h_2^2 \varphi(2, j), \ldots, h_2^2 \varphi(M-2, j),$$
$$h_2^2 \bar{\varphi}(M-1, j)), \; 1 \leqslant j \leqslant N-1,$$

donde

$$\bar{\varphi}(1, j) = \varphi(1, j) + \frac{1}{h_1^2} g(0, j) + \frac{h_1^2 + h_2^2}{12} g_{\bar{x}_2 x_2}(0, j)\Big),$$
$$\bar{\varphi}(M-1, j) = \varphi(M-1, j) +$$
$$+ \frac{1}{h_1^2}\left(g(M, j) + \frac{h_1^2 + h_2^2}{12} g_{\bar{x}_2 x_2}(M, j)\right)$$

y

$$F_j = (g(1, j), g(2, j), \ldots, g(M-1, j)), \quad j = 0, N.$$

En este caso las matrices B y A corresponden a operadores de diferencias Λ_1 y Λ, donde

$$\Lambda_1 y = y + \frac{h_1^2 + h_2^2}{12} y_{\bar{x}_1 x_1}, \quad h_1 \leqslant x_1 \leqslant l_1 - h_1,$$
$$\Lambda y = 2y - \frac{5h_2^2 - h_1^2}{6} y_{\bar{x}_1 x_1}, \quad h_1 \leqslant x_1 \leqslant l_1 - h_1,$$

y $y = 0$ para $x_1 = 0$ y $x_1 = l_1$. Estas matrices son tridiagonales, y, como no es difícil de comprobar, son conmutables.

El problema de contorno (29) se puede reducir al problema (1), (2). Para esto es necesario multiplicar el miembro primero de cada una de las ecuaciones (29) por B^{-1}, si existe una matriz inversa a B. Hallemos una condición suficiente para la existencia de B^{-1}. Es evidente, que la matriz inversa a B existirá, si el sistema de ecuaciones algebraicas lineales

$$BY = F \tag{30}$$

tiene solución única para cualquier miembro derecho F.

En virtud de la definición de la matriz B, (30) puede ser escrito en forma del esquema de diferencias.

$$\begin{gathered}\Lambda_1 y = y + \frac{h_1^2+h_2^2}{12} y_{\bar{x}_1 x_1} = f, \quad h_1 \leqslant x_1 \leqslant l_1 - h_1, \\ y(0) = y(l_1) = 0.\end{gathered} \tag{31}$$

En el § 1 del cap. II fue mostrado, que si para el esquema (31) se cumplen las condiciones suficientes de estabilidad del método de factorización, entonces la solución de la ecuación (31) existe y es única para cualquier segundo miembro f, y además ella puede ser hallada por el método de factorización. Anotando la derivada de diferencias $y_{\bar{x}_1 x_1}$ por puntos, escribamos (31) en forma de las ecuaciones escalares tripuntuales

$$\begin{gathered}-A_i y_{i-1} + C_i y_i - B_i y_{i+1} = F_i, \\ 1 \leqslant i \leqslant M-1, \\ y_0 = 0, \quad y_M = 0,\end{gathered} \tag{32}$$

donde $A_i = B_i \dfrac{h_1^2+h_2^2}{12h_1^2}$, $C_i = \dfrac{h_1^2+h_2^2}{6h_1^2} - 1$.

Recordemos, que para (32) las condiciones suficientes de estabilidad del método de factorización poseen la forma $|C_i| \geqslant |A_i| + |B_i|$, $i = 1, 2, \ldots, M-1$. De estas condiciones hallaremos, que la matriz B posee la inversa, si los pasos de la red $\bar{\omega}$ satisfacen la acotación $h_2 \leqslant \sqrt{2} h_1$. Bajo el cumplimiento de esta condición el problema (29) puede ser reducido al problema (1), (2) con $C = B^{-1}A$.

§ 2. Método de reducción completa para el primer problema de contorno

1. Proceso de exclusión impar-par. Pasemos ahora a la descripción del método de reducción completa. Comencemos por el primer problema de contorno para las ecuaciones vectoriales tripuntuales

$$-Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1,$$
$$Y_0 = F_0, \quad Y_N = F_N. \tag{1}$$

La idea del método de reducción completa para resolver el problema (1) consiste en la eliminación consecutiva en las ecuaciones (1) de las incógnitas Y_j primeramente con números j impares, a continuación de las restantes ecuaciones se eliminan las incógnitas Y_j con números j múltiples de 2, después los múltiplos de 4 y así sucesivamente. Cada paso del proceso de exclusión disminuye el número de incógnitas, y si N es una potencia de 2, es decir $N = 2^n$, entonces como resultado del proceso de exclusión queda una ecuación de la cual se puede hallar $Y_{N/2}$. El paso inverso del método consiste en encontrar sucesivamente las incógnitas Y_j primeramente con números j múltiplos de $N/4$, después múltiplos de $N/8$, $N/16$ y así sucesivamente.

Es evidente, que el método de reducción completa es una modificación del método de reducción completa de Gauss, aplicado al problema (1), en el cual la eliminación de las incógnitas se lleva a efecto según un orden especial. Recordemos, que en diferencia de este método, el método de factorización matricial, la eliminación de las incógnitas se efectúa según un orden natural.

Así pues, sea $N = 2^n$, $n > 0$. Por comodidad introduciremos las siguientes notaciones: $C^{(0)} = C$, $F_j^{(0)} = F_j$, $j = 1, 2, \ldots, N-1$, con ayuda de las cuales escribiremos (1) en la forma

$$-Y_{j-1} + C^{(0)}Y_j - Y_{j+1} = F_j^{(0)}, \quad 1 \leqslant j \leqslant N-1,$$
$$N = 2^n, \tag{1'}$$
$$Y_0 = F_0, \quad Y_N = F_N.$$

Examinemos el primer paso del proceso de exclusión. En este paso de las ecuaciones del sistema (1′) para los j múltiplos de 2, eliminaremos las incógnitas Y_j con números j impares. Para eso escribamos tres ecuaciones (1′) que

van consecutivamente:

$$-Y_{j-2} + C^{(0)}Y_{j-1} - Y_j = F_{j-1}^{(0)}$$
$$-Y_{j-1} + C^{(0)}Y_j - Y_{j+1} = F_j^{(0)},$$
$$-Y_j + C^{(0)}Y_{j+1} - Y_{j+2} = F_{j+1}^{(0)},$$
$$j = 2, 4, 6, \ldots, N-2.$$

Multipliquemos la segunda ecuación a la izquierda por $C^{(0)}$ y sumemos las tres ecuaciones obtenidas. Como resultado tendremos

$$-Y_{j-2} + C^{(1)}Y_j - Y_{j+2} = F_j^{(1)},$$
$$j = 2, 4, 6, \ldots, N-2 \qquad (2)$$
$$Y_0 = F_0, \quad Y_N = F_N,$$

donde

$$C^{(1)} = [C^{(0)}]^2 - 2E,$$
$$F_j^{(1)} = F_{j-1}^{(0)} + C^{(0)}F_j^{(0)} + F_{j+1}^{(0)},$$
$$j = 2, 4, 6, \ldots, N-2.$$

El sistema (2) contiene las incógnitas Y_j solamente con números j pares, el número de incógnitas en (2) es igual a $N/2 - 1$, y si este sistema es resuelto, entonces en virtud de (1') las incógnitas Y_j con números impares pueden ser halladas de las ecuaciones,

$$C^{(0)}Y_j = F_j^{(0)} + Y_{j-1} + Y_{j+1},$$
$$j = 1, 3, 5, \ldots, N-1 \qquad (3)$$

con miembros derechos ya conocidos.

De esta forma, el problema inicial (1') es equivalente al sistema (2) y a las ecuaciones (3), además por su estructura el sistema (2) es análogo al sistema inicial.

En el segundo paso del proceso de exclusión de las ecuaciones del sistema reducido (2) para los j múltiplos de (4), se eliminan las incógnitas con números j múltiplos de 2, pero no múltiplos de 4. Por analogía con el primer paso se toman tres ecuaciones del sistema (2):

$$-Y_{j-4} + C^{(1)}Y_{j-2} - Y_j = F_{j-2}^{(1)},$$
$$-Y_{j-2} + C^{(1)}Y_j - Y_{j+2} = F_j^{(1)},$$
$$-Y_j + C^{(1)}Y_{j+2} - Y_{j+4} = F_{j+2}^{(1)}, \quad j = 4, 8, 12, \ldots, N-4,$$

la segunda ecuación se multiplica a la izquierda por $C^{(1)}$ y se suman las tres ecuaciones. Como resultado obtenemos

un sistema de $N/4 - 1$ ecuaciones, que contiene las incógnitas Y_j con números múltiplos de 4:

$$-Y_{j-4} + C^{(2)}Y_j - Y_{j+4} = F_j^{(2)}, \quad j = 4,\ 8,\ 12,\ \ldots,\ N-4,$$
$$Y_0 = F_0, \quad Y_N = F_N;$$

las ecuaciones $C^{(1)}Y_j = F_j^{(1)} + Y_{j-2} + Y_{j+2}$, $j = 2, 6, 10, \ldots, N - 2$, para encontrar las incógnitas con números múltiplos de 2 pero no múltiplos de 4, y las ecuaciones (3) para las incógnitas con números impares. A su vez la matriz $C^{(2)}$ y el miembro derecho $F_j^{(2)}$ se determinan por las fórmulas

$$C^{(2)} = [C^{(1)}]^2 - 2E,$$
$$F_j^{(2)} = F_{j-2}^{(1)} + C^{(1)}F_j^{(1)} + F_{j+2}^{(1)}, \quad j = 4,\ 8,\ 12,\ \ldots,\ N-4.$$

Este proceso de exclusión puede ser continuado. Como resultado del l-ésimo paso obtendremos un sistema reducido para las incógnitas con números múltiplos de 2^l:

$$-Y_{j-2^l} + C^{(l)}Y_j - Y_{j+2^l} = F_j^{(l)}, \quad j = 2^l,\ 2\cdot 2^l,$$
$$3\cdot 2^l,\ \ldots,\ N - 2^l$$
$$Y_0 = F_0, \quad Y_N = F_N, \tag{4}$$

y los grupos de ecuaciones

$$C^{(k-1)}Y_j = F_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}},$$
$$j = 2^{k-1},\quad 3\cdot 2^{k-1},\quad 5\cdot 2^{k-1},\ \ldots,\ N - 2^{k-1}, \tag{5}$$

resolviendo las cuales sucesivamente para $k = l,\ l - 1, \ldots$ $\ldots, 1$, hallaremos las incógnitas restantes. Las matrices $C^{(k)}$ y los miembros derechos $F_j^{(k)}$ se encuentran por las fórmulas recurrentes

$$C^{(k)} = [C^{(k-1)}]^2 - 2E,$$
$$F_j^{(k)} = F_{j-2^{k-1}}^{(k-1)} + C^{(k-1)}F_j^{(k-1)} + F_{j+2^{k-1}}^{(k-1)}, \tag{6}$$
$$j = 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k,\ \ldots,\ N - 2^k,$$

para $k = 1, 2, \ldots$

De (4) se deduce, que después del $(n - 1)$-ésimo paso de exclusión ($l = n - 1$) queda una ecuación para $Y_{2^{n-1}} = Y_{N/2}$:

$$C^{(n-1)}Y_j = F_j^{(n-1)} + Y_{j-2^{n-1}} + Y_{j+2^{n-1}} =$$
$$= F_j^{(n-1)} + Y_0 + Y_N, \quad j = 2^{n-1},$$
$$Y_0 = F_0, \quad Y_N = F_N$$

con miembro derecho conocido. Uniendo esta ecuación con (5), obtendremos, que todas las incógnitas se encuentran consecutivamente de las ecuaciones

$$C^{(k-1)}Y_j = F_j^{(k-1)} + Y_{j-2^{k-1}} - Y_{j+2^{k-1}}, \qquad Y_0 = F_0, \quad Y_N = F_N, \tag{7}$$
$$j = 2^{k-1},\ 3\cdot 2^{k-1},\ 5\cdot 2^{k-1}, \ldots, N - 2^{k-1}, \qquad k = n,\ n-1, \ldots, 1.$$

De esta forma, las fórmulas (6) y (7) describen totalmente el método de reducción completa. Por las fórmulas (6) se transforman los miembros derechos, y de las ecuaciones (7) se encuentra la solución del problema inicial (1).

Al método descrito lo llamaremos método de reducción completa, ya que aquí en el sistema se realiza hasta el final la disminución sucesiva del número de ecuaciones, mientras que quede una ecuación para $Y_{N/2}$. En el método de reducción incompleta que será examinado en el capítulo IV, se realiza solamente una disminución parcial del orden del sistema y el sistema «reducido» se resuelve por un método especial.

2. Transformación del segundo miembro e inversión de las matrices. El cómputo del miembro derecho **F** por las fórmulas recurrentes (6) puede conducir a la acumulación de errores de cálculo, si la norma de la matriz $C^{(k-1)}$ es mayor que la unidad. Además, las matrices $C^{(k)}$ son, en general, matrices completas, aún cuando la matriz inicial $C^{(0)} = C$ sea tridiagonal. Y esto influye de manera esencial en el aumento del volumen de trabajo computacional para el cálculo de $F_j^{(k)}$ según las fórmulas (6). Para los ejemplos examinados en el § 1 la norma de la matriz, realmente de forma significativa, supera la unidad y tal algoritmo del método será computacionalmente inestable.

Para evitar esta dificultad, calcularemos en lugar de los vectores $F_j^{(k)}$ los vectores $p_j^{(k)}$, los cuales están relacionados con $F_j^{(k)}$ por las siguientes relaciones:

$$F_j^{(k)} \equiv \sum_{l=0}^{k-1} C^{(l)} p_j^{(k)} 2^{(k)}, \tag{8}$$

al mismo tiempo pondremos formalmente $\prod_{l=0}^{-1} C^{(l)} = E$, así que $p_j^{(0)} \equiv F_j^{(0)} \equiv F_j$.

Hallemos las relaciones recurrentes, a las que satisfacen los $p_j^{(k)}$. Para eso sustituyamos (8) en (6). Considerando, que

$C^{(l)}$ es la matriz no degenerada para cualquier l, de (6) obtendremos

$$2\prod_{l=0}^{k-1} C^{(l)} p_j^{(k)} = \prod_{l=0}^{k-2} C^{(l)} \left[p_{j-2^{k-1}}^{(k-1)} + C^{k-1} p_j^{(k-1)} + p_{j+2^{k-1}}^{(k-1)}\right]$$

$$2C^{k-1}p_j^{(k)} = p_{j-2^{k-1}}^{(k-1)} + C^{(k-1)} p_j^{k-1} + p_{j+2^{k-1}}^{(k-1)}. \qquad (9)$$

Designando $S_j^{(k-1)} = 2p_j^{(k)} - p_j^{(k-1)}$, obtendremos de (9), que los $p_j^{(k)}$ pueden ser hallados sucesivamente por las siguientes fórmulas:

$$C^{(k-1)}S_j^{(k-1)} = p_{j-2^{k-1}}^{(k-1)} + p_{j+2^{k-1}}^{(k-1)}, \quad p_j^{(k)} = 0{,}5\,(p_j^{(k-1)} + S_j^{(k-1)}),$$
$$j = 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k, \ldots, N-2^k, \quad k = 1, 2, \ldots, n-1,$$
$$p_j^{(0)} = F_j. \qquad (10)$$

Las relaciones recurrentes (10) contienen suma de vectores, producto de un vector por un número e inversión de las matrices $C^{(k-1)}$.

Queda ahora por eliminar $F_j^{(k-1)}$ de las ecuaciones (7). Sustituyendo (8) en (7), obtenemos

$$C^{(k-1)}Y_j = 2^{k-1}\prod_{l=0}^{k-2} C^{(l)} p_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}},$$
$$Y_0 = F_0, \quad Y_N = F_N, \qquad (11)$$
$$j = 2^{k-1},\ 3\cdot 2^{k-1}, \ldots, N-2^{k-1}, \quad k = n,\ n-1, \ldots, 1.$$

Aquí también es necesario invertir las matrices $C^{(k-1)}$, pero, además, en el miembro derecho de (11) apareció el producto de una matriz por un vector. En el algoritmo examinado más abajo el método utilizado para invertir las matrices $C^{(k-1)}$ permite evitar la operación no deseada de multiplicar una matriz por un vector y a su vez reducir la realización de (11) a la inversión de matrices y suma de vectores.

Examinemos ahora el problema de la inversión de las matrices C^{k-1}, definidas por las fórmulas recurrentes (6)

$$C^{(k)} = [C^{(k-1)}]^2 - 2E, \; k = 1, 2, \ldots, \; C^{(0)} = C. \qquad (12)$$

De (12) se deduce, que $C^{(k)}$ es un polinomio matricial de grado $2^{(k)}$ con respecto a C y con coeficiente unidad en la mayor potencia. Este polinomio se expresa mediante los conocidos polinomios de Chebishev de la siguiente forma:

$$C^{(k)} = 2T_{2^k}\left(\frac{1}{2}C\right), \quad k = 0, 1, \ldots, \qquad (13)$$

donde $T_n(x)$ es el polinomio de Chebishev de n-ésimo grado y primer género (véase el punto 2 del § 4 del cap. 1):

$$T_n(x) = \begin{cases} \cos(n \arccos x), & |x| \leqslant 1, \\ \frac{1}{2}[(x+\sqrt{x^2-1})^n + (x+\sqrt{x^2-1})^{-n}], & |x| \geqslant 1. \end{cases}$$

En efecto, en virtud de las propiedades del polinomio $T_n(x)$

$$T_{2n}(x) = 2[T_n(x)]^2 - 1, \quad T_1(x) = x,$$

de (12) se deduce (13) de manera evidente.

A continuación, utilizando la relación

$$\prod_{l=0}^{k-2} 2T_{2^l}(x) = U_{2^{k-1}-1}(x),$$

que conecta los polinomios de Chebishev de primer género con los polinomios de segundo género $U_n(x)$, donde

$$U_n(x) = \begin{cases} \dfrac{\operatorname{sen}((n+1)\arccos x)}{\operatorname{sen}(\arccos x)}, & |x| \leqslant 1, \\ \dfrac{1}{2\sqrt{x^2-1}}[(x+\sqrt{x^2-1})^{n+1} - (x+\sqrt{x^2-1})^{-(n+1)}] \\ |x| \geqslant 1, \end{cases}$$

es fácil calcular el producto de los polinomios $C^{(l)}$

$$\prod_{l=0}^{k-2} C^{(l)} = U_{2^{k-1}-1}\left(\frac{1}{2}C\right). \tag{14}$$

Así hemos obtenido una expresión explícita para $C^{(k)}$ y $\prod_{l=0}^{k-1} C^{(l)}$.

En lo sucesivo nos será necesario el lema 6 (véase el punto 5 del § 4 del cap. II). De acuerdo con el lema 6 cualquier cociente $g_m(x)/f_n(x)$ de polinomios sin raíces comunes en el caso $n > m$ y de raíces simples de $f_n(x)$, se desarrolla de la siguiente forma en fracciones elementales:

$$\frac{g_m(x)}{f_n(x)} = \sum_{l=1}^{n} \frac{a_l}{x-x_l}, \quad a_l = \frac{g_m(x_l)}{f'_n(x_l)},$$

donde x_l son las raíces del polinomio $f_n(x)$.

Utilicemos el lema 6 para el desarrollo de los cocientes $1/T_n(x)$ y $U_{n-1}(x)/T_n(x)$ en fracciones elementales. Las raíces del polinomio $T_n(x)$ son conocidas:

$$x_l = \cos\frac{(2l-1)\pi}{2n}, \quad l = 1, 2, \ldots, n, \tag{15}$$

y en estos puntos el polinomio $U_{n-1}(x)$ toma los valores distintos de cero

$$U_{n-1}(x_l) = \frac{\operatorname{sen}(\arccos x_l)}{\operatorname{sen}(\arccos x_l)} = \frac{(-1)^{l+1}}{\operatorname{sen}\frac{(2l-1)}{2n}\pi}, \quad l = 1, 2, \ldots, n.$$

Por eso, utilizando la relación $T'_n(x) = nU_{n-1}(x)$, del lema 6 obtenemos los siguientes desarrollos:

$$\frac{1}{T_n(x)} = \sum_{l=1}^{n} \frac{(-1)^{l+1}\operatorname{sen}\frac{(2l-1)\pi}{2n}}{n(x-x_l)}, \tag{16}$$

$$\frac{U_{n-1}(x)}{T_n(x)} = \sum_{l=1}^{n} \frac{1}{n(x-x_l)}, \tag{17}$$

donde x_l está definido en (15). Los desarrollos necesitados han sido hallados.

Obtengamos ahora las expresiones para las matrices $[C^{(k-1)}]^{-1}$ y $[C^{(k-1)}]^{-1}\prod_{l=0}^{k-2} C^{(l)}$ mediante la matriz C. De (13) y (14) teniendo en cuenta los desarrollos de los polinomios algebraicos (16) y (17), obtendremos

$$[C^{(k-1)}]^{-1} = \sum_{l=1}^{2^{k-1}} \alpha_{l,k-1}\left(C - 2\cos\frac{(2l-1)\pi}{2^k}E\right)^{-1},$$

$$[C^{(k-1)}]^{-1}\prod_{l=0}^{k-2} C^{(l)} = \frac{1}{2^{k-1}}\sum_{l=1}^{2^{k-1}}\left(C - 2\cos\frac{(2l-1)\pi}{2^k}E\right)^{-1}.$$

Las relaciones halladas permiten escribir en la siguiente forma tanto las fórmulas (10):

$$S_j^{(k-1)} + \sum_{l=1}^{2^{k-1}} \alpha_{l,k-1}C_{l,k-1}^{-1}\left(p_{j-2^{k-1}}^{(k-1)} + p_{j+2^{k-1}}^{(k-1)}\right),$$

$$p_j^{(k)} = 0,5\left(p_j^{(k-1)} + S_j^{(k-1)}\right), \tag{18}$$

$$p_j^{(0)} = F_j,$$
$$j = 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k, \ldots, N-2^k, \quad k = 1, 2, \ldots, n-1,$$

como las fórmulas (11):

$$Y_j = \sum_{l=1}^{2^{k-1}} C_{l,\,k-1}^{-1}\,[p^{(k-1)} + \alpha_{l,\,k-1}(Y_{j-2^{k-1}} + Y_{j+2^{k-1}})],$$
$$Y_0 = F_0, \quad Y_N = F_N, \tag{19}$$
$$j = 2^{k-1},\ 3\cdot 2^{k-1},\ 5\cdot 2^{k-1}, \ldots, N-2^{k-1}, \quad k = n, n-1, \ldots, 1,$$

donde hemos designado

$$C_{l,\,k-1} = C - 2\cos\frac{(2l-1)\pi}{2^k}E, \quad \alpha_{l,\,k-1} =$$
$$= \frac{(-1)^{l+1}}{2^{k-1}}\operatorname{sen}\frac{(2l-1)\pi}{2^k}. \tag{20}$$

Así, han sido obtenidas las fórmulas transformadas (18) y (19), que describen el método de reducción completa para la resolución del problema (1). Estas fórmulas contienen solamente operaciones de suma de vectores, de multplicación de un vector por un número y de inversión de matrices.

Notemos, que si C es una matriz tridiagonal, entonces será también tridiagonal toda matriz $C_{l,\,k-1}$. El problema de invertir tales matrices fue resuelto en el capítulo II. Sucesivamente, si para la matriz C se cumple la condición $(C, Y, Y) \geqslant 2\,(Y, Y)$, entonces de (20) se deduce, que las matrices $C_{l,\,k}$ serán definidas positivas y, por consiguiente, tendrán inversas acotadas. Entonces del desarrollo de $[C^{(k-1)}]^{-1}$ obtendremos que para todo $k \geqslant 1$ las matrices $C^{(k-1)}$ no son degeneradas. Recordemos, que esta suposición se utilizó para obtener las fórmulas (10).

3. **Algoritmo del método.** Las fórmulas (18) y (19) obtenidas más arriba sirven de base para el primer algoritmo del método. Examinemos ante todo, cuales valores intermedios y en cual etapa deben calcularse y recordarse para su posterior utilización.

El análisis de las fórmulas (19) muestra, que si k es fijo para el cálculo de Y_j se utilizan vectores $p_j^{(k-1)}$ con números $j = 2^{k-1},\ 3\cdot 2^{k-1}, \ldots, N-2^{k-1}$. Cualquier vector $p_j^{(l)}$ con el número j, pero con número l menor que $k-1$, es auxiliar y se guarda en la memoria provisionalmente. Por eso los vectores $p_j^{(k)}$ definidos en el k-ésimo paso por (18) pueden

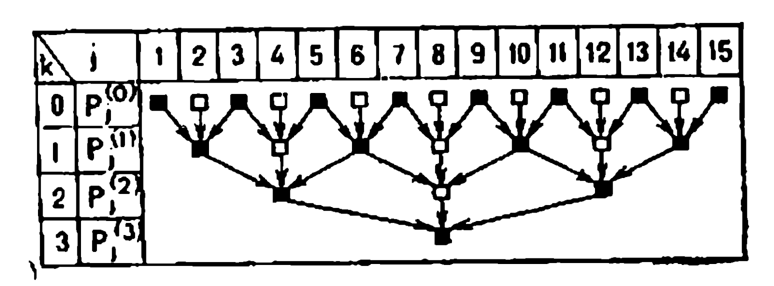


Fig. 1.

distribuirse en el lugar de $p_j^{(k-1)}$, al igual que las incógnitas Y_j, calculadas según (19). El método no exige memoria complementaria de ordenador, todos los vectores $p_j^{(k)}$ se distribuyen en el mismo lugar donde después se distribuirán los Y_j.

Ilustremos sobre un ejemplo la organización de los cálculos en el algoritmo examinado. Sea $N = 16$ ($n = 4$). En la fig. 1 se muestra la sucesión del cálculo y la recordación de los vectores $p_j^{(k)}$. Un cuadrado sombreado significa, que para el valor indicado del índice k, se guarda en la memoria para su posterior utilización el vector $p_j^{(k)}$ con el número j correspondiente. Respectivamente un cuadrado no sombreado significa, que $p_j^{(k)}$ es auxiliar y se recuerda provisionalmente en el lugar indicado. Las flechas indican cuales vectores $p_j^{(k-1)}$ se utilizan para calcular $p_j^{(k)}$.

Como resultado del paso directo del método serán recordados los siguientes vectores $p_j^{(k)}$:

$$p_1^{(0)},\ p_2^{(1)},\ p_3^{(0)},\ p_4^{(2)},\ p_5^{(0)},\ p_6^{(1)},\ p_7^{(0)},\ p_8^{(3)},$$
$$p_9^{(0)},\ p_{10}^{(1)},\ p_{11}^{(0)},\ p_{12}^{(2)},\ p_{13}^{(0)},\ p_{14}^{(1)},\ p_{15}^{(0)}.$$

Ellos se utilizan para calcular Y_j en el paso inverso del método.

En la fig. 2 se muestra la sucesión del cálculo de las incógnitas Y_j (notación simbólica ○). Mediante flechas se indica cuales Y_j son halladas en los pasos anteriores y cuales $p_j^{(k-1)}$ (notación simbólica ■) se utilizan para calcular Y_j si k es prefijado.

Pasemos ahora a la descripción del algoritmo del método de reducción completa. De acuerdo con (18), el paso directo del método se realiza de la siguiente forma:

1) Se dan los valores para $p_j^{(0)} = F_j$, $j = 1, 2, \ldots$ $\ldots, N-1$.

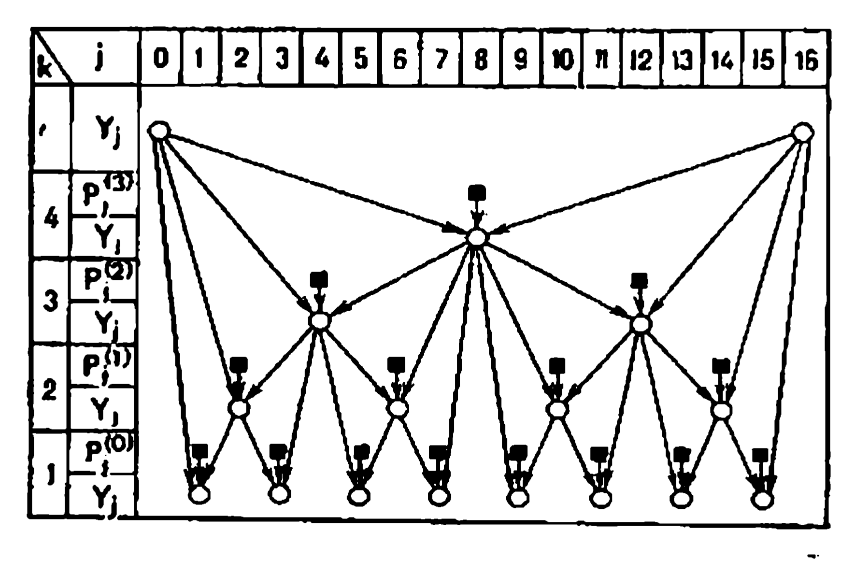


Fig. 2.

2) Para cada $k = 1, 2, \ldots, n - 1$ fijo y con $j = 2^k$, $2 \cdot 2^k, \ldots, N - 2^k$ también fijo, primeramente se calculan y se recuerdan los vectores

$$\varphi = p^{(k-1)}_{j-2^{k-1}} + p^{(k-1)}_{j+2^{k-1}}. \tag{21}$$

A continuación se resuelven las ecuaciones

$$C_{l,\, k-1} v_l = \alpha_{l,\, k-1} \varphi \tag{22}$$

para $l = 1, 2, \ldots, 2^{-1}$.
Como producto de la acumulación gradual del resultado en el lugar de $p_j^{(k-1)}$ se encuentra $p_j^{(k)}$

$$p_j^{(\ )} = 0{,}5\,(p_j^{(k-1)} + v_1 + v_2 + \ldots + v_{2^{k-1}}). \tag{23}$$

De acuerdo con (19), el paso inverso del método se realiza de la siguiente manera:

1) Se dan los valores para Y_0 y Y_N: $Y_0 = F_0$ y $Y_N = F_N$.

2) Para cada $k = n, n - 1, \ldots, 1$ fijo y con $j = 2^{k-1}$, $3 \cdot 2^{k-1}$, $5 \cdot 2^{k-1}, \ldots, N - 2^{k-1}$ fijo se calculan y se recuerdan los vectores

$$\varphi = Y_{j-2^{k-1}} + Y_{j+2^{k-1}}, \quad \psi = p_j^{k-1}. \tag{24}$$

Después se resuelven las ecuaciones

$$C_{l,\, k-1} v_l = \psi + \alpha_{l,\, k-1} \varphi, \tag{25}$$

para $l = 1, 2, \ldots, 2^{k-1}$.

Como resultado de la acumulación gradual de valores

en el lugar de $p_j^{(k-1)}$ se encuentra el vector de las incógnitas Y_j

$$Y_j = v_1 + v_2 + \ldots + v_{2(k-1)}. \qquad (26)$$

Contemos ahora el número de operaciones aritméticas gastadas en la realización del algoritmo descrito. Sea M la dimensión del vector de las incógnitas Y_j y designemos mediante $\mathring{q}$ el número de operaciones exigidas para resolver una ecuación del tipo (22) ó (25) para un miembro derecho dado. Consideraremos, que las magnitudes $\alpha_{l,\,k}$ son dadas de antemano.

Contemos primeramente el número Q_1 de operaciones, gastadas en el paso directo. En el cálculo del vector φ por las fórmulas (21) para k y j fijos se exigen M operaciones. Más adelante, en el cálculo del miembro derecho en (22) y en la resolución de la ecuación (22) se exigen $M + \mathring{q}$ operaciones. Por eso para encontrar todos los v_l se necesitan $2^{k-1}(M + \mathring{q})$ operaciones. El cálculo de $p_j^{(k)}$ por la fórmula (23) se realiza con un gasto de $2^{k-1}M + M$ operaciones. De esta forma, para calcular $p_j^{(k)}$ para un k y j es necesario gastar $M + 2^{k-1}(2M + \mathring{q})$ operaciones.

A continuación, para cada k fijo se necesitan calcular $N/2^k - 1$ diferentes $p_j^{(k)}$. Por consiguiente, la cantidad total Q_1 de operaciones, gastadas en la realización del paso directo, es igual a

$$Q_1 = \sum_{k=1}^{n-1} [M + (2M + \mathring{q})\, 2^{k-1}] \left(\frac{N}{2^k} - 1\right) =$$

$$= (M + 0{,}5\mathring{q})\, Nn - (M + \mathring{q})\, N = M(n-1) + \mathring{q}. \qquad (27)$$

Contemos ahora el número Q_2 de operaciones gastadas en el paso inverso. Para k y j fijos se exigen M operaciones en el cálculo según las fórmulas (24), $(2M + \mathring{q})\, 2^{k-1}$ operaciones para encontrar todos los v_l en la (25) y $(2^{k-1} - 1)\, M$ operaciones en el cálculo de los Y_j mediante la fórmula (26). Como el número de diferentes valores de j, para los cuales se realizan los cálculos indicados con un k fijo, es igual a $N/2^k$, entonces Q_2 es igual a

$$Q_2 = \sum_{k=1}^{n} [M + (2M + \mathring{q})\, 2^{k-1} + (2^{k-1} - 1)\, M]\, \frac{N}{2^k} =$$

$$= (1{,}5M + 0{,}5\mathring{q})\, Nn. \qquad (28)$$

Sumando (27) y (28) y teniendo en cuenta que $n = \log_2 N$, obtenemos la siguiente estimación para el número de operaciones del método de reducción completa, realizado según el algoritmo citado más arriba

$$Q = Q_1 + Q_2 = (2{,}5M + \mathring{q}) N \log_2 N - (M + \mathring{q}) N - M(n-1) + \mathring{q}. \qquad (29)$$

De (29) se deduce, que si $\mathring{q} = O(M)$, entonces $Q = O(MN \log_2 N)$.

4. **Segundo algoritmo del método.** El mérito principal del algoritmo construido es la exigencia mínima a la memoria del ordenador ya que no exige memoria complementaria para conservar la información auxiliar. Esta cualidad se alcanza al precio de cierto aumento del volúmen de trabajo computacional, el cual se gasta en el cálculo reiterado de las magnitudes intermedias. Examinemos otro algoritmo del método, el cual se caracteriza por un menor volúmen de trabajo computacional, pero que exige memoria complementaria, comparable en magnitud con el número total de incógnitas en el problema.

Para la construcción del segundo algoritmo regresemos a las fórmulas (6) y (7) que desriben el método de reducción completa:

$$C^{(k)} = [C^{(k-1)}]^2 - 2E,$$

$$F_j^{(k)} = F_{j-2^{k-1}}^{(k-1)} + C^{(k-1)} F_j^{(k-1)} + F_{j+2^{k-1}}^{(k-1)}, \qquad (6')$$

$$j = 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k,\ \ldots,\ N - 2^k,\quad k = 1, 2, \ldots, n-1,$$

$$C^{(k-1)} Y_j = F_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}},$$

$$Y_0 = F_0,\quad Y_N = Y_N \qquad (7')$$

$$j = 2^{k-1},\ 3\cdot 2^{k-1},\ 5\cdot 2^{k-1},\ \ldots,\ N - 2^{k-1},\quad k = n, n-1, \ldots, 1.$$

Aquí, como en el primer algoritmo, los vectores $F_j^{(k)}$ no se calculan directamente, y en lugar de ellos se determinan los vectores $p_j^{(k)}$ y $q_j^{(k)}$ relacionados con $F_j^{(k)}$ por la siguiente relación:

$$F_j^{(k)} = C^{(k)} p_j^{(k)} + q_j^{(k)}, \qquad (30)$$

$$j = 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k,\ \ldots,\ N - 2^k,$$

$$k = 0, 1, \ldots, n-1.$$

Hallemos las fórmulas recurrentes para calcular los vectores $p_j^{(k)}$ y $q_j^{(k)}$. Ya que en lugar de un vector $F_j^{(k)}$ nosotros introdujimos dos vectores, entonces se tiene una determinada arbitrariedad en la definición de $p_j^{(k)}$ y $q_j^{(k)}$. Elijamos $p_j^{(0)}$ y $q_j^{(0)}$ de manera tal, que se satisfaga la condición inicial $F_j^{(0)} \equiv F_j$. Para eso pongamos

$$p_j^{(0)} = 0, \quad q_j^{(0)} = F_j, \quad j = 1, 2, \ldots, N-1. \tag{31}$$

A continuación, sustituyendo (30) en (6′), obtendremos

$$C^{(k)}p_j^{(k)} + q_j^{(k)} = C^{(k-1)}[q_j^{(k-1)} + p_{j-2^{k-1}}^{(k-1)} + $$
$$+ C^{(k-1)}p_j^{(k-1)} + p_{j+2^{k-1}}^{(k-1)}] + q_{j-2^{k-1}}^{(k-1)} + q_{j+2^{k-1}}^{(k-1)},$$
$$j = 2^k,\ 2\cdot 2^k, \ldots, N-2^k, \qquad k = 1, 2, \ldots, n-1.$$

Eligiendo

$$q_j^{(k)} = 2p_j^{(k)} + q_{j-2^{k-1}}^{(k-1)} + q_{j+2^{k-1}}^{(k-1)} \tag{32}$$

y teniendo en cuenta, que $C^{(k)} + 2E = [C^{(k-1)}]^2$, de aquí hallaremos

$$C^{(k-1)}p_j^{(k)} = q_j^{(k-1)} + p_{j-2^{k-1}}^{(k-1)} + C^{(k-1)}p_j^{(k-1)} + p_{j+2^{k-1}}^{(k-1)}. \tag{33}$$

Aquí de nuevo suponemos, que $C^{(l)}$ para cualquier l es una matriz no degenerada.

Designando $S_j^{(k-1)} = p_j^{(k)} - p_j^{(k-1)}$, obtendremos de (31) — — (33) las siguientes fórmulas recurrentes para el cálculo de los vectores $p_j^{(k)}$ y $q_j^{(k)}$:

$$C^{(k-1)}S_j^{(k-1)} = q_j^{(k-1)} + p_{j-2^{k-1}}^{(k-1)} + p_{j+2^{k-1}}^{(k-1)},$$
$$p_j^{(k)} = p_j^{(k-1)} + S_j^{(k-1)},$$
$$q_j^{(k)} = 2p_j^{(k)} + q_{j-2^{k-1}}^{(k-1)} + q_{j+2^{k-1}}^{(k-1)}, \tag{34}$$
$$q_j^{(0)} \equiv F_j, \quad p_j^{(0)} \equiv 0,$$
$$j = 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k, \ldots, N-2^k, \qquad k = 1, 2, \ldots, n = 1.$$

Queda por excluir $F_j^{(k-1)}$ de la fórmula (7′). Sustituyendo (30) en (7′) y designando $t_j^{(k-1)} = Y_j - p_j^{(k-1)}$

obtendremos las siguientes fórmulas para el cálculo de Y_j:

$$\begin{aligned} &C^{(k-1)}t_j^{(k-1)} = q_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}}, \\ &Y_j = p_j^{(k-1)} + t_j^{(k-1)}, \\ &Y_0 = F_0, \quad Y_N = F_N, \\ &j = 2^{k-1},\ 3\cdot 2^{k-1},\ 5\cdot 2^{k-1}, \ldots, N - 2^{k-1}, \\ &\qquad k = n,\ n-1, \ldots, 1. \end{aligned} \tag{35}$$

De esta forma, obtuvimos las fórmulas (34) y (35) en las cuales se basa el segundo algoritmo del método de reducción completa. Estas fórmulas contienen las operaciones de suma de vectores e inversión de las matrices $C^{(k-1)}$.

Detengámonos ahora en el problema de invertir las matrices $C^{(k-1)}$. Como fue mostrado más arriba, la matriz $C^{(k)}$ es un polinomio de grado 2^k con respecto a la matriz inicial C y se determina por la fórmula (13) mediante el polinomio de Chebishev de primer género $T_n(x)$:

$$C^{(k)} = 2T_{2^k}\left(\frac{1}{2}C\right),$$

además el coeficiente del término de mayor grado es igual a la unidad. Como las raíces del polinomio $T_n(x)$ son conocidas (véase (15)), entonces $C^{(k)}$ se puede representar en la siguiente forma factorizada:

$$C^{(k)} = \prod_{l=1}^{2^k}\left(C - 2\cos\frac{(2l-1)\pi}{2^{k+1}}E\right), \qquad k = 0, 1, \ldots$$

Utilizando las notaciones (20), se puede escribir la matriz $C^{(k-1)}$ en la siguiente forma:

$$C^{(k-1)} = \prod_{l=1}^{2^{k-1}} C_{l,k-1}, \quad C_{l,k-1} = C - 2\cos\frac{(2l-1)\pi}{2^k}E. \tag{36}$$

La factorización (36) permite resolver fácilmente las ecuaciones del tipo $C^{(k-1)}v = \varphi$ con la parte derecha φ prefijada. El siguiente algoritmo da la solución de este problema mediante la inversión sucesiva de los factores en (36):

$$v_0 = \varphi, \quad C_{l,k-1}v_l = v_{l-1}, \quad l = 1, 2, \ldots, 2^{k-1},$$

además $v = v_{2^{k-1}}$. Nosotros utilizaremos este algoritmo para invertir las matrices $C^{(k-1)}$.

Describamos ahora el segundo algoritmo del método de reducción completa. El paso directo del método se realiza a base de (34) de la siguiente forma:

1) Se dan los valores $q_j^{(0)}$: $q_j^{(0)} = F_j$, $j = 1, 2, \ldots, N-1$.

2) El primer paso, para $k = 1$, se lleva a efecto por separado según fórmula que tienen en cuenta los datos iniciales $p_j^{(0)} \equiv 0$. Se resuelven las ecuaciones para $p_j^{(1)}$ y se calculan los $q_j^{(1)}$:

$$C p_j^{(1)} = q_j^{(0)}, \tag{37}$$
$$q_j^{(1)} = 2p_j^{(1)} + q_{j-1}^{(0)} + q_{j+1}^{(0)}, \quad j = 2, 4, 6, \ldots, N-2.$$

3) Para cada $k = 2, 3, \ldots, n-1$ fijo se calculan y se guardan en la memoria los vectores

$$v_j^{(0)} = q_j^{(k-1)} + p_{j-2^{k-1}}^{(k-1)} + p_{j+2^{k-1}}^{(k-1)},$$
$$j = 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k, \ldots, N-2^k. \tag{38}$$

A continuación para $l = 1, 2, 3, \ldots, 2^{k-1}$ fijo y para cada $j = 2^k, 2\cdot 2^k, 3\cdot 2^k, \ldots, N-2^k$ se resuelven las ecuaciones

$$C_{l,k-1} v_j^{(l)} = v_j^{(l-1)} \tag{39}$$

con una misma matriz, pero con segundos miembros diferentes. Como resultado serán hallados los vectores $v_j^{(2^{k-1})}$ (en las fórmulas (34) a estos vectores corresponden $S_j^{(k-1)}$). Los vectores $p_j^{(k)}$ y $q_j^{(k)}$ se calculan por las fórmulas

$$p_j^{(k)} = p_j^{(k-1)} + v_j^{(2^{k-1})},$$
$$q_j^{(k)} = 2p_j^{(k)} + q_{j-2^{k-1}}^{(k-1)} + q_{j+2^{k-1}}^{(k-1)}, \tag{40}$$
$$j = 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k, \ldots, N-2^k.$$

El paso inverso del método se realiza de acuerdo a (35):

1) Se prefijan los valores para Y_0 y Y_N: $Y_0 = F_0$ y $Y_N = F_N$.

2) Para cada $k = n, n-1, \ldots, 2$ fijo se calculan y se guardan en la memoria los vectores

$$v_j^{(0)} = q_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}},$$
$$j = 2^{k-1},\ 3\cdot 2^{k-1},\ 5\cdot 2^{k-1}, \ldots, N-2^{k-1}. \tag{41}$$

Después para $l = 1, 2, \ldots, 2^{k-1}$ fijo y para cada $j = 2^{k-1}, 3\cdot 2^{k-1}, 5\cdot 2^{k-1}, \ldots, N-2^{k-1}$ se resuelven las

ecuaciones

$$C_{l,\,k-1}v_j^{(l)} = v_j^{(l-1)}. \tag{42}$$

Como resultado se encuentran los vectores $v_j^{(2^{k-1})}$ (en (35) a ellos corresponden los vectores $t_j^{(k-1)}$). Seguidamente se calcula Y_j por la fórmula

$$Y_j = p_j^{(k-1)} + v_j^{(2^{k-1})},$$
$$j = 2^{k-1},\ 3\cdot 2^{k-1},\ 5\cdot 2^{k-1},\ \ldots,\ N-2^{k-1}. \tag{43}$$

3) La última operación del paso inverso para $k = 1$ se lleva a efecto con la solución de la ecuación

$$CY_j = q_j^{(0)} + Y_{j-1} + Y_{j+1},\quad j = 1, 3, 5, \ldots, N-1. \tag{44}$$

OBSERVACION AL ALGORITMO. Todos los vectores $p_j^{(k)}$ recientemente definidos por las fórmulas (37) y (40) se sitúan en el lugar de $p_j^{(k-1)}$. Todos los vectores $v_j^{(l)}$ en las fórmulas (38), (39), (41), (42) definidos hace poco por las fórmulas (37) y (40), los vectores $q_j^{(k)}$ o igualmente la solución Y_j de (43) y (44) se distribuyen en el lugar de $q_j^{(k-1)}$. Por consiguiente este algoritmo exige memoria del ordenador en 1,5 veces más, que el número de incógnitas en el problema.

La disminución del volúmen de trabajo computacional en el algoritmo dado en comparación con el primer algoritmo se basa en que al resolver las series de problemas (39) y (42) para diferentes j con iguales matrices $C_{l,\,k-1}$ el volúmen total del trabajo se gasta en resolver solamente el primer problema de la serie, y al resolver cada problema subsiguiente ya se exige significativamente menos operaciones aritméticas. Citemos el número de operaciones para el segundo algoritmo designando, como antes, mediante $\mathring{q}$ el número de operaciones gastadas en la resolución de una ecuación del tipo (39) o (42) para el segundo miembro prefijado, y mediante $\bar{q}$ el número de operaciones para resolver la misma ecuación pero con otro segundo miembro ($\bar{q} < \mathring{q}$).

El número de operaciones gastadas en la realización del paso directo es igual a

$$Q_1 = \sum_{k=1}^{n-1} \left\{6M\left(\frac{N}{2^k} - 1\right) + \left[\mathring{q} + \bar{q}\left(\frac{N}{2^k} - 2\right)\right]2^{k-1}\right\} -$$

$$- 3M\left(\frac{N}{2} - 1\right) = 0{,}5\bar{q}Nn + (0{,}5\mathring{q} - 1{,}5\bar{q} + 4{,}5M)N - 6Mn - (\mathring{q} - 2\bar{q} + 3M),$$

y en la realización del paso inverso

$$Q_2 = \sum_{k=1}^{n}\left\{3M\frac{N}{2^k} + \left[\mathring{q} + \left(\frac{N}{2^k} - 1\right)\bar{q}\right]2^{k-1}\right\} - \frac{MN}{2} = 0{,}5\bar{q}Nn + (\mathring{q} - \bar{q} + 2{,}5M)N - \mathring{q} + \bar{q} - 3M.$$

El número total de operaciones para el segundo algoritmo es igual a

$$Q = Q_1 + Q_2 = \bar{q}N\log_2 N + (1{,}5\mathring{q} - 2{,}5\bar{q} + 7M)N - 6Mn - 2\mathring{q} + 3\bar{q} - 6M. \quad (45)$$

De la estimación (45) se deduce, que si $\mathring{q} = O(M)$, entonces $\bar{q} = O(M)$ y $Q = O(MN\log_2 N)$, además aquí el coeficiente del término principal $MN\log_2 N$ es menor que en la estimación (29), ya que $\bar{q} < \mathring{q}$.

Detengámonos brevemente en otra singularidad del segundo algoritmo. Si bien en el primer algoritmo la inversión de las matrices $C^{(k-1)}$ se ha realizado simultáneamente por inversión de los factores $C_{l,\,k-1}$ y la subsiguiente sumación de los resultados, en el segundo algoritmo ocurre una inversión sucesiva de los factores y el resultado se obtiene después de invertir el último factor. Desde el punto de vista del proceso computacional real, el cual tiene en cuenta los errores de redondeo, el orden de inversión de los factores $C_{l,\,k-1}$ en el segundo algoritmo es esencial. Con una situación análoga nos encontraremos en el capítulo VI al estudiar el método iterativo de Chebishev.

Se puede recomendar el siguiente orden de inversión de las matrices $C_{l,\,k-1}$. A la matriz $C^{(k-1)}$ le ponemos en correspondencia el vector $\theta_{2^{k-1}}$ de dimensión 2^{k-1}, cuyas componentes son los números enteros de 1 hasta 2^{k-1}. Sea

$$\theta_{2^{k-1}} = \{\theta_{2^{k-1}}(1), \theta_{2^{k-1}}(2), \ldots, \theta_{2^{k-1}}(2^{k-1})\},$$

es decir, el l-ésimo elemento del vector $\theta_{2^{k-1}}$ se denota mediante $\theta_{2^{k-1}}(l)$. El número $\theta_{2^{k-1}}(l)$ define el turno de inversión de la matriz $C_{l,\,k-1}$.

El vector $\theta_{2^{k-1}}$ se construye recurrentemente. Sea $\theta_2 = \{2, 1\}$. Entonces el proceso de duplicación de la dimen-

sión del vector se describe por las siguientes fórmulas:

$$\theta_{2m}=\{\theta_{2m}(4i-3)=\theta_m(2i-1),$$
$$\theta_{2m}(4i-2)=\theta_m(2i-1)+m,$$
$$\theta_{2m}(4i-1)=\theta_m(2i)+m,\quad \theta_{2m}(4i)=\theta_m(2i),$$
$$i=1,2,\ldots,m/2\},\quad m=2,4,8,\ldots$$

Ejemplo: $\theta_{16}=\{2, 10, 14, 6, 8, 16, 12, 4, 3, 11, 15, 7, 5, 13, 9, 1\}$ y, por consiguiente, la matriz $C_{6,16}$ se invertirá la décimosexta y la matriz $C_{12,16}$ la séptima.

§ 3. Ejemplos de aplicación del método

1. **Problema de Dirichlet de diferencias para la ecuación de Poisson en un rectángulo.** Examinemos una aplicación del método de reducción completa construido más arriba a encontrar la solución del problema de Dirichlet de diferencias para la ecuación de Poisson en un rectángulo. Como fue mostrado antes, el problema de diferencias

$$y_{\bar{x}_1x_1}+y_{\bar{x}_2x_2}=-\varphi(x),\quad x\in\omega,$$
$$y(x)=g(x),\quad x\in\gamma,$$

dado sobre la red rectangular $\bar{\omega}=\{x_{ij}=(ih_1, jh_2),\ 0\leqslant i\leqslant M,\ 0\leqslant j\leqslant N,\ h_1M=l_1,\ h_2N=l_2\}$, se escribe en la forma del primer problema de contorno para las ecuaciones vectoriales tripuntuales

$$\begin{gathered}-Y_{j-1}+CY_j-Y_{j+1}=F_j,\quad 1\leqslant j\leqslant N-1,\\ Y_0=F_0,\quad Y_N=F_N.\end{gathered}\tag{1}$$

Aquí

$$Y_j=(y(1,j),\ y(2,j),\ \ldots,\ y(M-1,j)),$$
$$0\leqslant j\leqslant N,$$

es el vector de las incógnitas cuyas componentes son los valores de la función reticular $y(i, j)$ sobre la j-ésima fila de la red,

$$F_j=(h_2^2\varphi(1,j),\ h_2^2\varphi(2,j),\ \ldots,\ h_2^2\varphi(M-2,j),$$
$$h_2^2\bar{\varphi}(M-1,j)),\quad 1\leqslant j\leqslant N-1,$$
$$F_j=(g(1,j),\ g(2,j),\ \ldots,\ g(M-1,j)),\quad j=0,N,$$

donde

$$\bar{\varphi}(1, j) = \varphi(1, j) + \frac{1}{h_1^2} g(0, j),$$

$$\bar{\varphi}(M-1, j) = \varphi(M-1, j) + \frac{1}{h_1^2} g(M, j).$$

La matriz cuadrada C corresponde al operador de diferencias Λ, donde

$$\Lambda y = 2y - h_2^2 y_{\bar{x}_1 x_1}, \qquad h_1 \leqslant x \leqslant l_1 - h_1,$$
$$y = 0, \; x_1 = 0, \; l_1,$$

así que

$$CY_j = (\Lambda y(1, j), \Lambda y(2, j), \ldots, \Lambda y(M-1, j)).$$

El problema (1) puede ser resuelto por cualquiera de los dos algoritmos del método de reducción completa citados más arriba. La etapa fundamental de estos algoritmos es la resolución de las ecuaciones del tipo

$$C_{l, k-1} V = F, \; C_{l, k-1} = C - 2\cos\frac{(2l-1)\pi}{2^k} E \tag{2}$$

con el segundo miembro F dado. Aquí V es el vector de las incógnitas, $V = (v(1), v(2), \ldots, v(M-1))$ de la dimensión $M-1$ (para simplificar la escritura hemos omitido el índice en V y F).

Recordemos, que el número de operaciones gastadas en la resolución del problema (1) por el primer algoritmo, se determina mediante el número de operaciones $\mathring{q}$ exigidas para resolver la ecuación (2) (véase (29) del punto 3, § 2), y por el segundo algoritmo se determina fundamentalmente por medio del número de operaciones complementarias $\bar{q}$ que es necesario gastar para resolver la ecuación (2), pero con otro miembro derecho (véase (45) del punto 4, § 2).

Para el ejemplo examinado citemos el método de resolución de la ecuación (2) y estimemos $\mathring{q}$ y $\bar{q}$. De la definición de la matriz C se deduce, que la resolución de la ecuación (2) es equivalente a encontrar la solución del siguiente problema de diferencias:

$$2\left(1 - \cos\frac{(2l-1)\pi}{2^k}\right) v - h_2^2 v_{\bar{x}_1 x_1} = f(i),$$
$$1 \leqslant i \leqslant M-1, \qquad v(0) = v(M) = 0, \tag{3}$$

donde $f(i) = f_i$ es la i-ésima componente del vector F.

Anotando la derivada de diferencias $v_{\bar{x}_1 x_1}$ por puntos, escribamos (3) en la forma de una ecuación en diferencias tripuntual usual para las incógnitas escalares $v(i) = v_i$:

$$-v_{i-1} + av_i - v_{i+1} = bf_i, \quad 1 \leqslant i \leqslant M-1, \qquad v_0 = v_M = 0, \tag{4}$$

donde $a = 2\left[1 + b\left(1 - \cos\frac{(2l-1)\pi}{2^k}\right)\right]$, $b = \frac{h_2^2}{h_1^2}$. El problema (4) es un caso especial de los problemas de contorno tripuntuales, los métodos de resolución de los cuales fueron estudiados en el capítulo II. Fue mostrado, que un método efectivo de resolución de los problemas del tipo (4) es el método de factorización. Citemos las fórmulas de cálculo del método de factorización para el problema (4):

$$\alpha_{i+1} = 1/(a - \alpha_i), \quad i = 1, 2, \ldots, M-1, \quad \alpha_1 = 0,$$
$$\beta_{i+1} = (bf_i + \beta_i)\,\alpha_{i+1}, \quad i = 1, 2, \ldots, M-1, \; \beta_1 = 0,$$
$$v_i = \alpha_{i+1} v_{i+1} + \beta_{i+1}, \quad i = M-1, M-2, \ldots, 1,$$
$$v_M = 0.$$

De estas fórmulas se deduce, que el problema (4), y, por consiguiente, la ecuación (2), para a y b prefijadas, pueden ser resueltos con un gasto de $\overset{\circ}{q} = 7(M-1)$ operaciones. Para resolver la ecuación (2) con otro segundo miembro F no es necesario volver a contar los coeficientes de factorización α_i, y por eso el número complementario de operaciones $\bar{q}$ es igual a $\bar{q} = 5(M-1)$. Estas operaciones serán gastadas en el cálculo de β_i y en encontrar la solución v_i. Notemos, que el método de factorización para (4) será numéricamente estable, ya que se cumple la condición suficiente de estabilidad del método a los errores de redondeo, la cual en el caso dado posee la forma $a \geqslant 2$.

Sustituyendo en la estimación (29) del punto 3 del § 2 para el número de operaciones del primer algoritmo, obtendremos, manteniendo los términos principales, $Q^{(1)} \approx \approx 9.5\, MN \log_2 N - 8MN$. Para el segundo algoritmo de la estimación (45) del punto 4 del § 2, obtendremos la siguiente estimación para el número de operaciones: $Q^{2} \approx \approx 5MN \log_2 N + 5MN$. De esta forma, para cada uno de los algoritmos examinados el número de operaciones del método de reducción completa, aplicado para resolver el problema de Dirichlet de diferencias para la ecuación de

Poisson en un rectángulo, es una magnitud del orden $O(MN\log_2 N)$, y además para el segundo algoritmo se exigen menos operaciones aritméticas. Por ejemplo, para $M = N = 64$ obtendremos $Q^{(1)} \approx 1{,}4Q^{(2)}$ y para $M = N = 128$, $Q^{(1)} \approx 1{,}46Q^{(2)}$ respectivamente.

Nosotros no citaremos las fórmulas de cálculo para los algoritmos de resolución del problema de diferencias indicado, ya que en un nivel vectorial ellas están descritas detalladamente en el § 2.

En el punto 2 del § 1 fueron citados ejemplos de otros problemas de contorno de diferencias que se reducen al problema (1). Ellos se diferencian del problema de Dirichlet examinado por el tipo de las condiciones de contorno en los lados del rectángulo para $x_1 = 0$ y $x_1 = l_1$, lo cual conduce a matrices C diferentes. Así para el problema (10)—(12) del punto 2 del § 1, con condiciones de contorno de tercero o de segundo género para $x_1 = 0$, l_1, la ecuación (2) es equivalente al problema de diferencias

$$2\left(1-\cos\frac{(2l-1)\pi}{2k}\right)v - h_2^2 v_{\bar{x}_1x_1} = f, \qquad 1\leqslant i\leqslant M-1,$$

$$2\left(1+\frac{h_2^2}{h_1}\varkappa_{-1}-\cos\frac{(2l-1)\pi}{2k}\right)v - \frac{2h_2^2}{h_1}v_{x_1} = f, \qquad i=0,$$

$$2\left(1+\frac{h_2^2}{h_1}\varkappa_{+1}-\cos\frac{(2l-1)\pi}{2k}\right)v + \frac{2h_2^2}{h_1}v_{\bar{x}_1} = f, \qquad i=M.$$

Este problema en la forma tripuntual usual tiene el aspecto

$$\begin{aligned} &-v_{i-1}+av_i-v_{i+1}=bf_i, \qquad 1\leqslant i\leqslant M-1,\\ &-v_0=\bar{\varkappa}_1v_1+\mu_1,\\ &\quad v_M=\bar{\varkappa}_2v_{M-1}+\mu_2, \end{aligned} \tag{5}$$

donde

$$\bar{\varkappa}_1=\frac{2}{a+2h_1\varkappa_{-1}}, \qquad \bar{\varkappa}_2=\frac{2}{a+2h_1\varkappa_{+1}},$$

$$\mu_1=\frac{bf_0}{a+2h_1\varkappa_{-1}}, \qquad \mu_2=\frac{bf_M}{a+2h_1\varkappa_{+1}},$$

siendo a y b definidos más arriba.

Ya que $a > 2$, $\varkappa_{\pm1}\geqslant 0$, entonces $0<\bar{\varkappa}_1<1$ y $0<\bar{\varkappa}_2<1$, el método de factorización para resolver el problema (5) será también estable, y en este caso el algoritmo del método de reducción completa exigirá $O(MN\log_2 N)$

operaciones aritméticas.

2. Problema de Dirichlet de diferencias con orden de exactitud aumentado. En el punto 4 del § 1 el problema de Dirichlet de diferencias para la ecuación de Poisson con orden de exactitud aumentado

$$Y_{\bar{x}_1 x_1} + Y_{\bar{x}_2 x_2} + \frac{h_1^2 + h_2^2}{12} Y_{\bar{x}_1 x_1 \bar{x}_2 x_2} = -\varphi(x), \quad x \in \omega,$$

$$y(x) = g(x), \quad x \in \gamma$$

se redujo al primer problema de contorno para la ecuación vectorial tripuntual no simplificada

$$\begin{gathered} -BY_{j-1} + AY_j - BY_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1, \\ Y_0 = F_0, \quad Y_N = F_N. \end{gathered} \tag{6}$$

Las matrices cuadradas B y A de la dimensión $(M-1) \times \times (M-1)$ corresponden a los operadores de diferencias Λ_1 y Λ, donde

$$\Lambda_1 y = y + \frac{h_1^2 + h_2^2}{12} y_{\bar{x}_1 x_1}, \quad h_1 \leqslant x_1 \leqslant l_1 - h_1,$$

$$\Lambda y = 2y - \frac{5h_2^2 - h_1^2}{6} y_{\bar{x}_1 x_1}, \quad h_1 \leqslant x_1 \leqslant l_1 - h_1$$

y $y = 0$ para $x_1 = 0$ y $x_1 = l_1$.

Se mostró, que si se cumple la condición $h_2 \leqslant \sqrt{2} h_1$, entonces la ecuación (6) se reduce a la forma ordinaria

$$\begin{gathered} -Y_{j-1} + CY_j - Y_{j+1} = \Phi_j, \quad 1 \leqslant j \leqslant N-1 \\ Y_0 = \Phi_0, \quad Y_N = \Phi_N, \end{gathered} \tag{7}$$

donde $C = B^{-1}A$, $\Phi_j = B^{-1}F_j$, $1 \leqslant j \leqslant N-1$ y $\Phi_j = = F_j$ para $j = 0, N$. Además, se observó, que las matrices A y B conmutan.

Para resolver (7) utilicemos el primer algoritmo del método. Ya que la matriz $C_{l,\,k-1}$ se puede escribir en la forma

$$C_{l,\,k-1} = C - 2\cos\frac{(2l-1)\pi}{2^k} E = B^{-1}\left(A - 2\cos\frac{(2l-1)\pi}{2^k} B\right),$$

entonces las fórmulas (18) y (19) del § 2 que definen el primer algoritmo, adquieren el siguiente aspecto:

$$S_j^{(k-1)} = \sum_{l=1}^{2^{k-1}} \alpha_{l,\,k-1}\left(A - 2\cos\frac{(2l-1)\pi}{2^k}B\right)^{-1} \times$$
$$\times B\left(p_{j-2^{k-1}}^{(k-1)} + p_{j+2^{k-1}}^{(k-1)}\right),$$
$$p_j^{(k)} = 0{,}5\left(p_j^{(k-1)} + S_j^{(k-1)}\right),$$
$$j = 2^k,\ 2\cdot 2^k,\ \ldots,\ N - 2^k, \qquad k = 1,\ 2,\ \ldots,\ n-1,$$
$$Bp_j^{(0)} \equiv F_j,$$
$$Y_j = \sum_{l=1}^{2^{k-1}} \left(A - 2\cos\frac{(2l-1)\pi}{2^k}B\right)^{-1} B\,[p_j^{(k-1)} +$$
$$+ \alpha_{l,\,k-1}(Y_{j-2^{k-1}} + Y_{j+2^{k-1}})],$$
$$Y_0 = F_0,\ Y_N = F_N, \quad j = 2^{k-1},\ 3\cdot 2^{k-1},\ \ldots,\ N - 2^{k-1},$$
$$k = n,\ n-1,\ \ldots,\ 1.$$

Para evitar la inversión de las matrices B siendo dado $p_j^{(0)}$ y la multiplicación de $p_j^{(k-1)}$ por la matriz B al calcular Y_j, haremos unos cambios, poniendo $\overline{p}_j^{(k)} = Bp_j^{(k)}$ y $\overline{S}_j^{(k)} = BS_j^{(k)}$. Entonces teniendo en cuenta la conmutatividad de las matrices A y B, y por consiguiente, de las matrices $\left(A - \cos\frac{(2l-1)\pi}{2^k}B\right)^{-1}$ y B, las fórmulas escritas más arriba tomarán la forma (se omite la raya encima de $\overline{p}_j^{(k)}$ y $\overline{S}_j^{(k)}$):

$$S_j^{(k-1)} \sum_{l=1}^{2^{k-1}} \alpha_{l,\,k-1}\left(A - 2\cos\frac{(2l-1)\pi}{2^k}B\right)^{-1} \times$$
$$\times B\left(p_{j-2^{k-1}}^{(k-1)} + p_{j+2^{k-1}}^{(k-1)}\right),$$
$$p_j^{(k)} = 0{,}5\left(p_j^{(k-1)} + S_j^{(k-1)}\right), \qquad j = 2^k,\ 2\cdot 2^k,\ \ldots,\ N - 2^k,$$
$$k = 1,\ 2,\ \ldots,\ n-1,$$
$$p_j^{(0)} \equiv F_j,$$
$$Y_j = \sum_{l=1}^{2k-1} \left(A - 2\cos\frac{(2l-1)\pi}{2^k}B\right)^{-1} \times$$
$$\times\,[p_j^{(k-1)} + \alpha_{l,\,k-1}B\,(Y_{j-2^{k-1}} + Y_{j+2^{k-1}}]),$$

$$Y_0 = F_0,\ Y_N = F_N, \qquad j = 2^{k-1},\ 3\cdot 2^{k-1}, \ldots, N-2^{k-1}$$
$$k = n,\ n-1, \ldots, 1.$$

Las fórmulas obtenidas generan la siguiente modificación en el primer algoritmo: la fórmula (21) del § 2 se sustituye por

$$\varphi = B\,(p^{(k-1)}_{j-2^{k-1}} + p^{(k-1)}_{j+2^{k-1}},$$

y en lugar de las ecuaciones (22) se resuelven las ecuaciones

$$\left(A - 2\cos\frac{(2l-1)\pi}{2^k}\,B\right) v_l = \alpha_{l,\,k-1}\varphi$$

con la φ calculada. Análogamente (24) se sustituye por

$$\varphi = B\,(Y_{j-2^{k-1}} + Y_{j+2^{k-1}}), \qquad \psi = p_j^{(k-1)},$$

y en lugar de (25) se resuelven las ecuaciones

$$\left(A - 2\cos\frac{(2l-1)\pi}{2^k}\,B\right) v_l = \psi + \alpha_{l,\,k-1}\varphi.$$

Por consiguiente, la etapa fundamental del algoritmo para el problema examinado es la resolución de ecuaciones del tipo

$$\left(A - 2\cos\frac{(2l-1)\pi}{2^k}\,B\right) V = F \tag{8}$$

con el segundo miembro F dado. Utilizando la definición de las matrices A y B con ayuda de los operadores de diferencias Λ y Λ_1 obtendremos, que (8) es equivalente a encontrar la solución del siguiente problema de diferencias:

$$2\left(1 - \cos\frac{(2l-1)\pi}{2^k}\right) v - \left(\frac{5h_2^2 - h_1^2}{6} + \right.$$
$$\left. + \frac{h_1^2 + h_2^2}{6}\cos\frac{(2l-1)\pi}{2^k}\right) v_{\bar{x}_1 x_1} = f, \tag{9}$$
$$1 \leqslant l \leqslant M-1, \qquad v_0 = v_M = 0.$$

Anotando esta ecuación por puntos, obtendremos el primer problema de contorno para la ecuación tripuntual escalar

$$\begin{aligned} &-v_{i-1} + a v_i - v_{i+1} = b f_i, \quad 1 \leqslant i \leqslant M-1, \\ &v_0 = v_M = 0, \end{aligned} \tag{10}$$

donde

$$a = 2\left[1 + b\left(1 - \cos\frac{(2l-1)\pi}{2k}\right)\right],$$

$$b = \frac{6h_1^2}{5h_2^2 - h_1^2 + (h_1^2 + h_2^2)\cos\frac{(2l-1)\pi}{2k}}.$$

El problema de diferencias (10) puede ser resuelto por el método de factorización, el cual será numéricamente estable, si se cumple la condición $|a| \geqslant 2$. Mostremos, que para cualesquiera h_1 y h_2 se cumple esta condición. En efecto, si h_1 y h_2 son tales, que se cumple la desigualdad

$$\frac{h_2^2}{h_1^2} \geqslant \frac{1 - \cos\frac{(2l-1)\pi}{2k}}{5 + \cos\frac{(2l-1)\pi}{2k}}, \tag{11}$$

entonces $0 < b \leqslant \infty$ y, por consiguiente, $a > 2$. Notemos, que siendo la igualdad en (11), el coeficiente de $v_{\bar{x}_1 x_2}$ en (9) se anula y v puede ser hallado de (9) por la fórmula explícita.

Si (11) no se cumple, entonces para b es cierta la estimación

$$b < -6/\left(1 - \cos\frac{(2l-1)\pi}{2k}\right),$$

y, por lo tanto, $a < -10$. La afirmación está demostrada.

De esta forma, para resolver el problema de Dirichlet de diferencias con orden de exactitud aumentado se puede aplicar el método de reducción completa con una estimación $O(MN \times \log_2 N)$ de operaciones aritméticas.

§ 4. Método de reducción completa para otros problemas de contorno

1. **Segundo problema de contorno.** Más arriba fue estudiado el método de reducción completa para resolver el primer problema de contorno para ecuaciones vectoriales tripuntuales. Comenzaremos el estudio del método para condiciones de contorno más complejas, examinando *el segundo problema de contorno*. Supongamos que se exige hallar la solución

del siguiente problema:

$$
\begin{aligned}
&CY_0 - 2Y_1 = F_0, \quad j = 0,\\
&-Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1, \qquad (1)\\
&-2Y_{N-1} + CY_N = F_N, \quad j = N,
\end{aligned}
$$

donde $N = 2^n$, $n > 0$.

El proceso de eliminación sucesiva de las incógnitas en (1) se realiza igual que en el caso de las condiciones de contorno de primer género. Precisamente, para los j pares tendremos las ecuaciones

$$
-Y_{j-2} + C^{(1)}Y_j - Y_{j+2} = F_j^{(1)}, \quad j = 2, 4, 6, \ldots, N-2, \qquad (2)
$$

y para los j impares, las ecuaciones

$$
C^{(0)}Y_j = F_j^{(0)} + Y_{j-1} + Y_{j+1}, \quad j = 1, 3, 5, \ldots, N-1, \qquad (3)
$$

donde, como antes, se utilizan las notaciones

$$
F_j^{(1)} = F_{j-1}^{(0)} + C^{(0)}F_j^{(0)} + F_{j+1}^{(0)}, \quad C^{(1)} = [C^{(0)}]^2 - 2E,
$$
$$
C^{(0)} = C, \quad F_j^{(0)} \equiv F_j.
$$

Quedan no transformadas las ecuaciones del sistema (1) solamente para $j = 0$ y $j = N$. Eliminemos de las ecuaciones indicadas las incógnitas Y_j con números impares j. Para eso utilicemos dos ecuaciones adyacentes. Escribamos las ecuaciones para $j = 0$ y $j = 1$:

$$
C^{(0)}Y_0 - 2Y_1 = F_0^{(0)}, \quad -Y_0 + C^{(0)}Y_1 - Y_2 = F_1^{(0)}
$$

Multipliquemos la primera ecuación por $C^{(0)}$ a la izquierda, y la segunda por 2, sumemos las ecuaciones obtenidas y hallaremos

$$
C^{(1)}Y_0 - 2Y_2 = F_0^{(1)}, \qquad (4)
$$

donde $F_0^{(1)} = C^{(0)}F_0^{(0)} + 2F_1^{(0)}$. Análogamente obtendremos la ecuación

$$
-2Y_{N-2} - 2 + C^{(1)}Y_N = F_N^{(1)}, \qquad (5)
$$

donde $F_N^{(1)} = 2F_{N-1}^{(0)} + C^{(0)}F_N^{(0)}$.

Uniendo (2), (4) y (5), obtendremos un sistema completo «reducido» de ecuaciones para las incógnitas con números j

pares, el cual tiene una estructura análoga a (1):

$$C^{(1)}Y_0 - 2Y_2 = F_0^{(1)}, \quad j = 0,$$
$$-Y_{j-2} + C^{(1)}Y_j - Y_{j+2} = F_j^{(1)}, \quad j = 2, 4, 6, \ldots, N-2,$$
$$-2Y_{N-2} + C^{(1)}Y_N = F_N^{(1)}, \quad j = N,$$

y un grupo de ecuaciones (3) para las incógnitas con números j impares.

Continuando más adelante el proceso descrito de eliminación de las incógnitas, después del n-ésimo paso de exclusión obtendremos el sistema para Y_0 y Y_N:

$$C^{(n)}Y_0 - 2Y_N = F_0^{(n)}, \quad -2Y_0 + C^{(n)}Y_N = F_N^{(n)} \tag{6}$$

y las ecuaciones para determinar las restantes incógnitas:

$$C^{(k-1)}Y_j = F_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}},$$
$$j = 2^{k-1}, \ 3\cdot 2^{k-1}, \ 5\cdot 2^{k-1}, \ \ldots, \ N - 2^{k-1}, \tag{7}$$
$$k = n, \ n-1, \ \ldots, \ 1,$$

donde $F_j^{(k)}$ y $C^{(k)}$ se definen por recurrencia para $k = 1, 2, \ldots, n$:

$$F_0^{(k)} = C^{(k-1)}F_0^{(k-1)} + 2F_{2^{k-1}}^{(k-1)},$$
$$F_j^{(k)} = F_{j-2^{k-1}}^{(k-1)} + C^{(k-1)}Y_j^{k-1} + F_{j+2^{k-1}}^{(k-1)},$$
$$j = 2^k, \ 2\cdot 2^k, \ 3\cdot 3^k, \ \ldots, \ N - 2^k, \tag{8}$$
$$F_N^{(k)} = 2F_{N-2^{k-1}}^{(k-1)} + C^{(k-1)}F_N^{(k-1)},$$
$$C^{(k)} = [C^{(k-1)}]^2 - 2E.$$

Así, hay que resolver el sistema (6) y luego sucesivamente de las ecuaciones (7) hallar todas las restantes incógnitas.

Aquí, como en el segundo algoritmo del método de reducción completa, aplicable al caso del primer problema de contorno, en lugar de los vectores $F_j^{(k)}$ determinaremos los vectores $p_j^{(k)}$ y $q_j^{(k)}$ enlazados con $F_j^{(k)}$ por la relación

$$F_j^{(k)} = C^{(k)}p_j^{(k)} + q_j^{(k)},$$
$$j = 0, \ 2^k, \ 2\cdot 2^k, \ 3\cdot 2^k, \ \ldots, \ N - 2^k, \ N, \tag{9}$$
$$k = 0, \ 1, \ \ldots, \ n.$$

De (8) hallaremos, como antes, que $p_j^{(k)}$ y $q_j^{(k)}$ para $j \neq 0$, N pueden ser obtenidos por las fórmulas:

$$
\begin{aligned}
&C^{(k-1)} S_j^{(k-1)} = q_j^{(k-1)} + p_{j-2^{k-1}}^{(k-1)} + p_{j+2^{k-1}}^{(k-1)}, \\
&p_j^{(k)} = p_j^{(k-1)} + S_j^{(k-1)}, \\
&q_j^{(k)} = 2p_j^{(k)} + q_{j-2^{k-1}}^{(k-1)} + q_{j+2^{k-1}}^{(k-1)}, \\
&j = 2^k,\ 2\cdot 2^k,\ \ldots,\ N-2^k, \quad k = 1,\ 2,\ \ldots,\ n-1, \\
&q_j^{(0)} \equiv F_j, \quad p_j^{(0)} \equiv 0.
\end{aligned} \tag{10}
$$

Hallemos ahora las fórmulas para $p_j^{(k)}$ y $q_j^{(k)}$ con $j = 0$, N. Sustituyendo (9) para $j = 0$ en (8), obtendremos para $F_0^{(k)}$

$$
C^{(k)} p_0^{(k)} + q_0^{(k)} = C^{(k-1)} [q_0^{(k-1)} + 2p_{2^{k-1}}^{(k-1)} + \\
+ C^{(k-1)} p_0^{(k-1)}] + 2q_{2^{k-1}}^{(k-1)}.
$$

Eligiendo $q_0^{(k)} = 2p_0^{(k)} + 2q_{2^{k-1}}^{(k-1)}$ y teniendo en cuenta la igualdad (12) del punto 1 § 2, encontramos la ecuación para $p_0^{(k)}$

$$
C^{(k-1)} p_0^{(k)} = C^{(k-1)} p_0^{(k-1)} + q_0^{(k-1)} + 2p_{2^{k-1}}^{(k-1)}.
$$

Así, los vectores $p_0^{(k)}$ y $q_0^{(k)}$ pueden ser hallados por las siguientes fórmulas recurrentes:

$$
\begin{aligned}
&C^{(k-1)} S_0^{(k-1)} = q_0^{(k-1)} + 2p_{2^{k-1}}^{(k-1)}, \\
&p_0^{(k)} = p_0^{(k-1)} + S_0^{(k-1)}, \\
&q_0^{(k)} = 2p_0^{(k)} + 2q_{2^{k-1}}^{(k-1)}, \quad k = 1,\ 2,\ \ldots,\ n, \\
&q_0^{(0)} = F_0, \quad p_0^{(0)} = 0.
\end{aligned} \tag{11}
$$

Las fórmulas para $p_N^{(k)}$ y $q_N^{(k)}$ se obtienen análogamente:

$$
\begin{aligned}
&C^{(k-1)} S_N^{(k-1)} = q_N^{(k-1)} + 2p_{N-2^{k-1}}^{(k-1)}, \\
&p_N^{(k)} = p_N^{(k-1)} + S_N^{(k-1)}, \\
&q_N^{(k)} = 2p_N^{(k)} + 2q_{N-2^{k-1}}^{(k-1)}, \quad k = 1,\ 2,\ \ldots,\ n, \\
&q_N^{(0)} = F_N, \quad p_N^{(0)} = 0.
\end{aligned} \tag{12}
$$

Así, las fórmulas (10)-(12) nos permiten hallar completamente todos los vectores $p_j^{(k)}$ y $q_j^{(k)}$ necesarios. Queda por

eliminar $F_j^{(h)}$ de (6) y (7). Sustituyendo (9) en (7), obtendremos las siguientes fórmulas para calcular los Y_j:

$$C^{(k-1)}t_j^{(k-1)} = q_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}},$$
$$Y_j = p_j^{(k-1)} + t_j^{(k-1)}, \tag{13}$$
$$j = 2^{k-1},\ 3\cdot 2^{k-1},\ 5\cdot 2^{k-1},\ \ldots,\ N-2^{k-1},$$
$$k = n,\ n-1,\ \ldots,\ 1.$$

Quedan por hallar Y_0 y Y_N de (6). Pero primero notemos, que de (11) y (12) para $k=n$ se deducen las igualdades

$$q_0^{(n)} = 2p_0^{(n)} + 2q_{2^{n-1}}^{(n-1)}, \quad q_N^{(n)} = 2p_N^{(n)} + 2q_{2^{n-1}}^{(n-1)},$$

es decir

$$q_0^{(n)} - q_N^{(n)} = 2\,(p_0^{(n)} - p_N^{(n)}). \tag{14}$$

A continuación de (9) y (14) obtendremos, que

$$F_0^{(n)} - F_N^{(n)} = C^{(n)}\,(p_0^{(n)} - p_N^{(n)}) + q_0^{(n)} - q_N^{(n)} =$$
$$= (C^{(n)} + 2E)\,(p_0^{(n)} - p_N^{(n)}).$$

Teniendo en cuenta la fórmula (12) del punto 1 § 2, tendremos definitivamente:

$$F_0^{(n)} - F_N^{(n)} = [C^{(n-1)}]^2\,(p_0^{(n)} - p_N^{(n)}). \tag{15}$$

Utilicemos las relaciones obtenidas para hallar Y_0 y Y_N de (6). Restando de la primera ecuación del sistema (6) la segunda, y teniendo en cuenta (15) y la igualdad (12) del punto 1 § 2, obtendremos, que

$$(C^{(n)} + 2E)\,(Y_0 - Y_N) = [C^{(n-1)}]^2\,(Y_0 - Y_N) =$$
$$= F_0^{(n)} - F_N^{(n)} = [C^{n-1}]^2\,(p_0^{(n)} - p_N^{(n)}).$$

Considerando, que $C^{(n-1)}$ es una matriz no degenerada, de aquí hallaremos

$$Y_0 = Y_N + p_0^{(n)} - p_N^{(n)}. \tag{16}$$

Sustituyendo el Y_0 hallado en la segunda ecuación del sistema (9), obtendremos la ecuación para encontrar Y_N:

$$B^{(n)}Y_N = F_N^{(n)} + 2\,(p_0^{(n)} - p_N^{(n)}) = B^{(n)}p_N^{(n)} + q_N^{(n)} + 2p_0^{(n)},$$

donde $B^{(n)} = C^{(n)} - 2E$. Por consiguiente, si designamos $t^{(n)} = Y_N - p_N^{(n)}$, entonces Y_N se puede hallar, resolviendo

la ecuación

$$B^{(n)}t^{(n)} = q_N^{(n)} + 2p_0^{(n)}, \quad (Y_N = p_N^{(n)} + t^{(n)}). \tag{17}$$

De (16) obtendremos, que Y_0 se puede hallar por la fórmula

$$Y_0 = p_0^{(n)} + t^{(n)}, \tag{18}$$

donde $t^{(n)}$ fue encontrado más arriba.

Así, las fórmulas (10)-(13), (17) y (18) describen el método de reducción completa para resolver el segundo problema de contorno para las ecuaciones vectoriales tripuntuales (1).

OBSERVACION. Si Y_0 es prefijado, es decir, si en lugar del problema (1) se resuelve el problema

$$-Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1,$$

$$-2Y_{N-1} + CY_N = F_N, \quad j = N, \quad Y_0 = F_0,$$

entonces no es necesario calcular los vectores $p_0^{(k)}$ y $q_0^{(k)}$, e Y_N, como se deduce de (6) y (9), se encuentra mediante la resolución de la ecuación

$$C^{(n)}t_N^{(n)} = q_N^{(n)} + 2Y_0, \quad (Y_N = p_N^{(n)} + t_N^{(n)}).$$

Análogamente, si se da Y_N, entonces no es necesario calcular los vectores $p_N^{(k)}$ y $q_N^{(k)}$, y Y_0 se determina de la ecuación $C^{(n)}t_0^{(n)} = q_0^{(n)} + 2Y_N$, $Y_0 = p_0^{(n)} + t_0^{(n)}$.

Para culminar la descripción del método de reducción es necesario indicar los procedimientos de inversión de las matrices $C^{(k)}$ y $B^{(n)} = C^{(n)} - 2E$. Para invertir las matrices $C^{(k-1)}$ se utiliza la factorización

$$C^{(k-1)} = \prod_{l=1}^{2^{k-1}} C_{l,\,k-1}, \quad C_{l,\,k-1} = C - 2\cos\frac{(2l-1)\pi}{2^k}E, \tag{19}$$

obtenida más arriba (véase (36) del § 2).

Notemos, que al cumplirse la condición $(CY, Y) \geqslant 2(Y, Y)$, todas las matrices $C_{l,\,k-1}$ son no degeneradas y, por lo tanto, es no degenerada la matriz $C^{(k-1)}$. Detengámonos más detalladamente en el problema de inversión de la matriz $B^{(n)}$.

De la definición de $B^{(n)}$ y de la relación (12) del punto 1 § 2, obtenemos

$$B^{(n)} = C^{(n)} - 2E = [C^{(n-1)}]^2 - 4E = (C^{(n-1)} + 2E)(C^{(n-1)} - 2E) =$$
$$= [C^{(n-2)}]^2 [C^{(n-1)} - 2E] = \dots$$
$$\dots = [C^{(n-2)} C^{(n-3)} \dots C^{(0)}]^2 (C^1 - 2E) =$$
$$= [C^{(n-2)} C^{(n-3)} \dots C^{(0)}]^2 (C^{(0)} - 2E)(C^{(0)} + 2E) =$$
$$= [\prod_{k=1}^{n-1} C^{(k-1)}]^2 (C - 2E)(C + 2E).$$

Sustituyendo aquí (19), hallaremos la siguiente representación para la matriz:

$$B^{(n)} = [\prod_{k=1}^{n-1} \prod_{l=1}^{2^{k-1}} C_{l,\,k-1}]^2 (C - 2E)(C + 2E). \qquad (20)$$

Así pues, la matriz $B^{(n)}$ ha sido factorizada y la inversión de $B^{(n)}$ puede ser realizada mediante la inversión consecutiva de los factores.

OBSERVACION 1. Se puede obtener una escritura más compacta de (20):

$$B^{(n)} = \prod_{l=1}^{n} \left(C - 2\cos\frac{l\pi}{2^{n-1}} E \right).$$

OBSERVACION 2. De (20) se deduce, que la matriz $B^{(n)}$ será no degenerada, si se cumple la condición $(CY, Y) > 2\,(Y, Y)$. Si existe un tal vector $Y^* \neq 0$, para el cual $CY^* = 2Y^*$, entonces $B^{(n)}$ es degenerada y es imposible la aplicación inmediata del método de reducción. Esto es una consecuencia de la degeneración de la matriz del sistema (1) en el caso examinado. Realmente, en este caso el sistema homogéneo (1) posee solución no nula $Y_j = Y^*$, y por eso el sistema (1) no es soluble para cualquier segundo miembro. Si para este segundo miembro existe la solución entonces ella no es única, y se determina con exactitud hasta el sumando Y^*. Una de las posibles soluciones se separa en la etapa de inversión de la matriz degenerada $B^{(n)}$. La situación indicada tiene lugar durante la resolución del problema de Neumann para la ecuación de Poisson en un rectángulo. Los problemas indicados serán examinados con más detalle en el capítulo XII, dedicado a la resolución de ecuaciones reticulares degeneradas.

2. Problema periódico. Los problemas vectoriales tripuntuales periódicos aparecen al resolver por métodos de diferencias las ecuaciones elípticas en sistemas de coordenadas curvilíneas ortogonales: sistemas cilíndricos, polares y esféricos. En el punto 3 del § 1 se citan ejemplos de problemas diferenciales, para los cuales los esquemas de diferencias pueden ser reducidos al siguiente problema: hallar la solución de las ecuaciones

$$\begin{gathered} -Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1, \\ -Y_{N-1} + CY_0 - Y_1 = F_0, \quad j = 0, \quad Y_N = Y_0. \end{gathered} \tag{21}$$

El problema (21) también puede ser resuelto por el método de reducción completa. Examinemos el primer paso del proceso de eliminación de las incógnitas. Como antes, de las ecuaciones del sistema (21) para $j = 2, 4, 6, \ldots, N-2$ excluiremos las incógnitas Y_j con números impares j por medio de dos ecuaciones contiguas. Obtendremos

$$-Y_{j-2} + C^{(1)}Y_j - Y_{j+2} = F_j^{(1)}, \quad j = 2, 4, 6, \ldots, N-2. \tag{22}$$

Queda por excluir Y_1 y Y_{N-1} de la ecuación (21) para $j = 0$. Para eso escribamos las siguientes tres ecuaciones del sistema (21):

$$\begin{gathered} -Y_0 + CY_1 - Y_2 = F_1, \quad j = 1, \\ -Y_{N-1} + CY_0 - Y_1 = F_0, \quad j = 0, \\ -Y_{N-2} + CY_{N-1} - Y_N = F_{N-1}, \quad j = N-1, \end{gathered}$$

multipliquemos la segunda ecuación a la izquierda por C, sumemos las tres ecuaciones y tengamos en cuenta que $Y_N = Y_0$. Como resultado obtendremos la ecuación

$$-Y_{N-2} + C^{(1)}Y_0 - Y_2 = F_0^{(1)}, \quad Y_N = Y_0, \tag{23}$$

donde

$$F_0^{(1)} = F_1^{(0)} + C^{(0)}F_0^{(0)} + F_{N-1}^{(0)}, \quad C^{(0)} = C, \quad F_j^{(0)} \equiv F_j.$$

Uniendo (22) y (23), obtenemos un sistema completo para las incógnitas Y_j con números j pares, el cual posee una estructura análoga a (21). Las incógnitas Y_j con números j impares se encuentran de las ecuaciones usuales

$$C^{(0)}Y_j = F_j^{(0)} + Y_{j-1} + Y_{j+1}, \quad j = 1, 3, 5, \ldots, N-1.$$

El proceso de exclusión puede ser continuado más adelante. Después del l-ésimo paso del proceso de exclusión obtendremos el sistema para las incógnitas Y_j con números j múltiplos de 2^l:

$$-Y_{j-2^l} + C^{(l)}Y_j - Y_{j+2^l} = F_j^l,$$
$$j = 2^l,\ 2\cdot 2^l,\ 3\cdot 2^l,\ \ldots,\ N-2^l,$$
$$-Y_{N-2^l} + C^{(l)}Y_0 - Y_{2^l} = F_0^{(l)}, \quad j=0, \quad Y_N = Y_0,$$

y el grupo de ecuaciones

$$C^{(k-1)}Y_j = F_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}},$$
$$j = 2^{k-1};\ 3\cdot 2^{k-1},\ 5\cdot 2^{k-1},\ \ldots,\ N-2^{k-1}, \quad k = l,\ l-1,\ \ldots,\ 1 \tag{24}$$

para encontrar sucesivamente las restantes incógnitas. Los segundos miembros $F_j^{(k)}$ se definen recurrentemente para $k = 1,\ 2,\ \ldots,\ n-1$:

$$F_j^{(k)} = F_{j-2^{k-1}}^{(k-1)} + C^{(k-1)}F_j^{(k-1)} + F_{j+2^{k-1}}^{(k-1)},$$
$$j = 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k,\ \ldots,\ N-2^k, \tag{25}$$
$$F_0^{(k)} F_{2^{k-1}}^{(k-1)} + C^{(k-1)}F_0^{(k-1)} + F_{N-2^{k-1}}^{(k-1)}, \quad F_j^{(0)} \equiv F_j.$$

Como resultado del $(n-1)$-ésimo paso del proceso de exclusión obtenemos un sistema respecto a Y_0 y $Y_{2^{n-1}}$ ($Y_N = Y_0$):

$$\begin{aligned} &C^{(n-1)}Y_0 - 2Y_{2^{n-1}} = F_0^{(n-1)}, \\ &-2Y_0 + C^{(n-1)}Y_{2^{n-1}} = F_{2^{n-1}}^{(n-1)}. \end{aligned} \tag{26}$$

Resolviendo este sistema, hallaremos Y_0, $Y_{2^{n-1}}$ y $Y_N = Y_0$, y en virtud de (24) las restantes incógnitas serán halladas como la solución de las ecuaciones

$$C^{(k-1)}Y_j = F_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}},$$
$$j = 2^{k-1},\ 3\cdot 2^{k-1},\ 5\cdot 2^{k-1},\ \ldots,\ N-2^{k-1},$$
$$k = n-1,\ n-2,\ \ldots,\ 1.$$

Antes de resolver (26), hallemos las fórmulas recurrentes para los vectores $p_j^{(k)}$ y $q_j^{(k)}$, relacionados con $F_j^{(k)}$ por la

siguiente razon:

$$F_j^{(k)} = C^{(k)} p_j^{(k)} + q_j^{(k)},$$
$$j = 0,\ 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k,\ \dots,\ N - 2^k.$$

Utilizando las fórmulas recurrentes (25) para $F_j^{(k)}$, obtenemos

$$\begin{aligned} &C^{(k-1)} S_j^{(k-1)} = q_j^{(k-1)} + p_{j-2^{k-1}}^{(k-1)} + p_{j+2^{k-1}}^{(k-1)},\\ &p_j^{(k)} = p_j^{(k-1)} + S_j^{(k-1)},\\ &q_j^{(k)} = 2p_j^{(k)} + q_{j-2^{k-1}}^{(k-1)} + q_{j+2^{k-1}}^{(k-1)}, \\ &j = 2^k,\ 2\cdot 2^k,\ 3\cdot 2^k,\ \dots,\ N - 2^k, \quad k = 1,\ 2,\ \dots,\ n-1,\\ &q_j^{(0)} \equiv F_j,\ \ p_j^{(0)} \equiv 0, \quad j = 1,\ 2,\ \dots,\ N-1, \end{aligned} \tag{27}$$

de las cuales se encuentran $p_j^{(k)}$ y $q_j^{(k)}$ para $j \neq 0$, y las fórmulas

$$\begin{aligned} &C^{(k-1)} S_0^{(k-1)} = q_0^{(k-1)} + p_{2^{k-1}}^{(k-1)} + p_{N-2^{k-1}}^{(k-1)},\\ &p_0^{(k)} = p_0^{(k-1)} + S_0^{(k-1)},\\ &q_0^{(k)} = 2p_0^{(k)} + q_{2^{k-1}}^{(k-1)} + q_{N-2^{k-1}}^{(k-1)},\ k = 1, 2,\ \dots,\ n-1, \end{aligned} \tag{28}$$

$$q_0^{(0)} = F_0, \quad p_0^{(0)} = 0$$

para encontrar $p_0^{(k)}$ y $q_0^{(k)}$.

Volvamos ahora a la resolución del sistema (26). De (27) y (28) para $k = n - 1$ obtenemos las relaciones

$$\begin{aligned} &q_{2^{n-1}}^{(n-1)} = 2p_{2^{n-1}}^{(n-1)} + q_{2^{n-2}}^{(n-2)} + q_{3\cdot 2^{n-2}}^{(n-2)},\\ &q_0^{(n-1)} = 2p_0^{(n-1)} + q_{2^{n-2}}^{(n-2)} + q_{3\cdot 2^{n-2}}^{(n-2)}, \end{aligned}$$

de las cuales hallamos

$$q_0^{(n-1)} - q_{2^{n-1}}^{(n-1)} = 2(p_0^{(n-1)} - p_{2^{n-1}}^{(n-1)}). \tag{29}$$

Restemos ahora de la primera ecuación del sistema (26) la segunda. Teniendo en cuenta (29) y la igualdad (12) del punto 1 del § 2, obtendremos

$$\begin{aligned} (C^{(n-1)} + 2E)(Y_0 - Y_{2^{n-1}}) &= \\ = [C^{(n-2)}]^2 (Y_0 - Y_{2^{n-1}}) &= F_0^{(n-1)} - F_{2^{n-1}}^{(n-1)} = \\ = C^{(n-1)} (p_0^{(n-1)} - p_{2^{n-1}}^{(n-1)}) &+ q_0^{(n-1)} - q_{2^{n-1}}^{(n-1)} = \\ &= [C^{(n-2)}]^2 (p_0^{(n-1)} - p_{2^{n-1}}^{(n-1)}). \end{aligned}$$

Suponiendo que $C^{(n-2)}$ es una matriz no degenerada, de aquí obtenemos la relación

$$Y_{2^{n-1}} = Y_0 - p_0^{(n-1)} + p_{2^{n-1}}^{(n-1)}. \tag{30}$$

Sustituyendo (30) en la primera ecuación del sistema (26), obtendremos

$$(C^{(n-1)} - 2E)\, Y_0 = F_0^{(n-1)} - 2\,(p_0^{(n-1)} - p_{2^{n-1}}^{(n-1)}) =$$
$$= (C^{(n-1)} - 2E)\, p_0^{(n-1)} + q_0^{(n-1)} + 2p_{2^{n-1}}^{(n-1)}.$$

Por consiguiente, Y_0 se puede hallar por las fórmulas

$$\begin{aligned} &B^{(n-1)} t^{(n-1)} = q_0^{(n-1)} + 2p_{2^{n-1}}^{(n-1)},\\ &B^{(n-1)} = C^{(n-1)} - 2E,\\ &Y_0 = p_0^{(n-1)} + t^{(n-1)}, \end{aligned} \tag{31}$$

y $Y_{2^{n-1}}$ en virtud de (30) se encuentra entonces de la relación

$$Y_{2^{n-1}} = p_{2^{n-1}}^{(n-1)} + t^{(n-1)}. \tag{32}$$

Las restantes incógnitas se hallan sucesivamente según las fórmulas

$$\begin{aligned} &Y_N = Y_0,\\ &C^{(k-1)} t_j^{(k-1)} = q_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}},\\ &Y_j = p_j^{(k-1)} + t_j^{(k-1)},\\ &j = 2^{k-1},\ 3\cdot 2^{k-1},\ 5\cdot 2^{k-1},\ \ldots,\ N - 2^{k-1},\\ &k = n-1,\ n-2,\ \ldots,\ 1. \end{aligned} \tag{33}$$

De esta manera, las fórmulas (27), (28), (31)-(33) describen el método de reducción completa para resolver el problema periódico (21). Para invertir las matrices $C^{(k-1)}$ y $B^{(n-1)}$ se utilizan las factorizaciones (19), (20) y además en (20) se necesita solamente cambiar n por $n-1$.

Citemos la valorización del número de operaciones aritméticas Q, que se exigen para la realización del método de reducción completa en el caso del problema periódico. Designemos, como antes, mediante $\mathring{q}$ el número de operaciones gastadas en la resolución de la ecuación $C_{l,\,k-1}V = F$, y mediante $\bar{q}$ el número de operaciones complementarias

para resolver esa misma ecuación, pero con otro miembro derecho F. La estimación se dá por la fórmula

$$Q = \bar{q} N \log_2 N + (1{,}5\mathring{q} - 2\bar{q} + 7M) N - 2\overset{3}{q} + \\ + 2\bar{q} - 14M.$$

La comparación de esta estimación con la estimación (45) del § 2, obtenido para el caso del primer problema de contorno, muestra que los gastos en la resolución del problema periódico son prácticamente iguales a los gastos en la resolución del primer problema de contorno.

3. Tercer problema de contorno

3.1 PROCESO DE ELIMINACION DE LAS INCOGNITAS. Examinemos ahora el método de reducción completa para resolver el tercer problema de contorno para ecuaciones vectoriales tripuntuales

$$\begin{aligned} &(C + 2\alpha E)\, Y_0 - 2Y_1 = F_0, \quad j = 0, \\ &-Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N - 1, \\ &-2Y_{N-1} + (C + 2\beta E)\, Y_N = F_N, \quad j = N. \end{aligned} \tag{34}$$

Suponiendo, que se cumplen las condiciones $\alpha \geqslant 0$, $\beta \geqslant 0$, $\alpha^2 + \beta^2 \neq 0$, introduzcamos las siguientes notaciones:

$$C^{(0)} = C, \quad C_1^{(0)} = C + 2\alpha E, \quad C_2^{(0)} = C + 2\beta E,$$
$$F_j^{(0)} \equiv F_j,$$

y utilizándolas escribamos (34) en la forma

$$\begin{aligned} &C_1^{(0)} Y_0 - 2Y_1 = F_0^{(0)}, \quad j = 0, \\ &-Y_{j-1} + C^{(0)} Y_j - Y_{j+1} = F_j^{(0)}, \quad 1 \leqslant j \leqslant N - 1 \\ &-2Y_{N-1} + C_2^{(0)} Y_N = F_N^{(0)}, \quad j = N. \end{aligned} \tag{34'}$$

Sea $N = 2^n$. El proceso de eliminación de las incógnitas para (34′) se realiza igual que para el sistema (1), el cual corresponde al caso $C_1^{(0)} = C_2^{(0)} = C^{(0)}$ $(\alpha = \beta = 0)$.

Escribamos el sistema reducido, obtenido como resultado del n-ésimo paso del proceso de eliminación de las incógnitas

$$C_1^{(n)} Y_0 - 2Y_N = F_0^{(n)}, \quad -2Y_0 + C_2^{(n)} Y_N = F_N^{(n)}, \tag{6'}$$

y los grupos de ecuaciones

$$C^{(k-1)}Y_j = F_j^{(k-1)} + Y_{j-2^{k-1}} + Y_{j+2^{k-1}}, \tag{35}$$
$$j = 2^{k-1},\ 3\cdot 2^{k-1},\ \ldots,\ N - 2^{k-1},$$
$$k = n,\ n-1,\ \ldots,\ 1$$

para encontrar sucesivamente las incógnitas Y_j. Aquí los miembros derechos $F_j^{(k)}$ se determinan por las fórmulas recurrentes:

$$F_j^{(k)} = F_{j-2^{k-1}}^{(k-1)} + C^{(k-1)}F_j^{(k-1)} + F_{j+2^{k-1}}^{(k-1)},$$
$$j = 2^k,\ 2\cdot 2^k,\ \ldots,\ N-2^k,\quad k=1,\ 2,\ \ldots,\ n-1, \tag{36}$$

$$F_0^{(k)} = C^{(k-1)}F_0^{(k-1)} + 2F_{2^{k-1}}^{(k-1)},\quad k=1,\ 2,\ \ldots,\ n, \tag{37}$$

$$F_N^{(k)} = 2F_{N-2^{k-1}}^{(k-1)} + C^{(k-1)}F_N^{(k-1)},\quad k=1,\ 2,\ \ldots,\ n, \tag{38}$$

y las matrices $C_1^{(k)}$, $C_2^{(k)}$ y $C^{(k)}$, por las fórmulas

$$C^{(k)} = [C^{(k-1)}]^2 - 2E,\quad k=1,\ 2,\ \ldots,\ n-1,\quad C^{(0)} = C,$$
$$C_1^{(k)} = C^{(k-1)}C_1^{(k-1)} - 2E,\quad k=1,\ 2,\ \ldots,\ n,\quad C_1^{(0)} = C + 2\alpha E, \tag{39}$$
$$C_2^{(k)} = C^{(k-1)}C_2^{(k-1)} - 2E,\quad k=1,\ 2,\ \ldots,\ n,\quad C_2^{(0)} = C + 2\beta E.$$

Del sistema (6′) obtenemos las ecuaciones para determinar Y_0 y Y_N. De (39) se puede obtener, que $C_1^{(k)}$, $C_2^{(k)}$ y $C^{(k)}$ son los polinomios matriciales de grado 2^k con respecto a una misma matriz C. Por consiguiente, ellas son conmutables. Por eso de (6′) obtenemos las ecuaciones

$$\mathscr{D}^{(n+1)}Y_0 = F_0^{(n+1)},\quad C_2^{(n)}Y_N = F_N^{(n)} + 2Y_0 \tag{40}$$

y las ecuaciones equivalentes a ellas

$$\mathscr{D}^{(n+1)}Y_N = F_N^{(n+1)},\quad C_1^{(n)}Y_0 = F_0^{(n)} + 2Y_N, \tag{40′}$$

donde se ha designado

$$F_0^{(n+1)} = C_2^{(n)}F_0^{(n)} + 2F_N^{(n)}, \tag{41}$$

$$F_N^{(n+1)} = 2F_0^{(n)} + C_1^{(n)}F_N^{(n)}, \tag{42}$$

$$\mathscr{D}^{(n+1)} = C_1^{(n)}C_2^{(n)} - 4E = C_2^{(n)}C_1^{(n)} - 4E. \tag{43}$$

De esta forma, para encontrar Y_0 y Y_N se puede hacer uso de las ecuaciones (40) ó (40′). Utilizaremos (40).

En lugar de los vectores $F_j^{(k)}$ determinaremos los vectores $p_j^{(k)}$ y $q_j^{(k)}$, que están relacionados con $F_j^{(k)}$ por las siguientes

expresiones:

$$F_0^{(k)} = C_1^{(k)} p_0^{(k)} + q_0^{(k)}, \tag{44}$$

$$F_N^{(k)} = C_2^{(k)} p_N^{(k)} + q_N^{(k)}, \quad k = 0,\ 1,\ \ldots,\ n, \tag{45}$$

$$F_0^{(n+1)} = \mathscr{D}^{(n+1)} p_0^{(n+1)} + q_0^{(n+1)}, \tag{46}$$

$$F_j^{(k)} = C^{(k)} p_j^{(k)} + q_c^{(k)}, \tag{47}$$

$$j = 2^k,\ 2 \cdot 2^k,\ \ldots,\ N - 2^k,\ k = 0,\ 1,\ 2,\ \ldots$$
$$\ldots,\ n - 1.$$

Obtengamos fórmulas recurrentes para $p_j^{(k)}$ y $q_j^{(k)}$. Si $j \neq 0,\ N$, entonces de (36), (39) y (47), suponiendo, como antes, la no degeneración de las matrices, obtenemos las fórmulas:

$$\begin{gathered} C^{(k-1)} S_j^{(k-1)} = q_j^{(k-1)} + p_{j-2^{k-1}}^{(k-1)} + p_{j+2^{k-1}}^{(k-1)}, \\ p_j^{(k)} = p_j^{(k-1)} + S_j^{(k-1)}, \\ q_j^{(k)} = 2 p_j^{(k-1)} + q_{j-2^{k-1}}^{(k-1)} + q_{j+2^{k-1}}^{(k-1)}, \\ j = 2^k,\ 2 \cdot 2^k,\ \ldots,\ N - 2^k, \quad k = 1,\ 2,\ \ldots,\ n - 1, \\ q_j^{(0)} \equiv F_j, \quad p_j^{(0)} \equiv 0. \end{gathered} \tag{48}$$

Hallemos las fórmulas para $p_0^{(k)}$ y $q_0^{(k)}$ siendo $k = 0,\ 1,\ \ldots,\ n + 1$. Sustituyendo (44) y (47) en (37), y (44)-(46) en (41), obtendremos para $k = 1,\ 2,\ \ldots,\ n$

$$C_1^{(k)} p_0^{(k)} + q_0^{(k)} = C^{(k-1)} (C_1^{(k-1)} p_0^{(k-1)} + q_0^{(k-1)} + 2 p_{2^{k-1}}^{(k-1)} + 2 q_{2^{k-1}}^{(k-1)} \tag{49}$$

y para $k = n + 1$

$$\mathscr{D}^{(n+1)} p_0^{(n+1)} + q_0^{(n+1)} = C_2^{(n)} (C_1^{(n)} p_0^{(n)} + q_0^{(n)} + 2 p_N^{(n)}) + 2 q_N^{(n)}. \tag{50}$$

Elijamos $q_0^{(k)}$ y $q_0^{(n+1)}$ según las fórmulas

$$\begin{gathered} q_0^{(k)} = 2 p_0^{(k)} + 2 q_{2^{k-1}}^{(k-1)}, \quad k = 1,\ 2,\ \ldots,\ n, \\ q_0^{(n+1)} = 4 p_0^{(n+1)} + 2 q_N^{(n)} \end{gathered} \tag{51}$$

y utilicemos las igualdades

$$C_1^{(k)} + 2E = C^{(k-1)} C_1^{(k-1)}, \quad \mathscr{D}^{(n+1)} + 4E = C_2^{(n)} C_1^{(n)},$$

que resultan de (39) y (43).

Entonces, bajo la condición de no degeneración de $C^{(k-1)}$ y $C_2^{(n)}$, (49) y (50) se pueden escribir en forma de la única ecuación

$$C_1^{(k-1)}p_0^{(k)}=C_1^{(k-1)}p_0^{(k-1)}+q_0^{(k-1)}+2p_{2^{k-1}}^{(k-1)},$$
$$k=1,\ 2,\ \ldots,\ n+1.$$

Uniendo estas ecuaciones con (51), obtenemos las fórmulas definitivas para calcular $p_0^{(k)}$ y $q_0^{(k)}$:

$$\begin{aligned}
&C_1^{(k-1)}S_0^{(k-1)}=q_0^{(k-1)}+2p_{2^{k-1}}^{(k-1)},\\
&p_0^{(k)}=p_0^{(k-1)}+S_0^{(k+1)},\quad k=1,\ 2,\ \ldots,\ n+1,\\
&q_0^{(k)}=2p_0^{(k)}+2q_{2^{k-1}}^{(k-1)},\quad k=1,\ 2,\ \ldots,\ n,\\
&q_0^{(n+1)}=4p_0^{(n+1)}+2q_N^{(n)},\\
&q_0^{(0)}=F_0,\quad p_0^{(0)}=0.
\end{aligned}\tag{52}$$

Análogamente, utilizando (45), (47) y las relaciones recurrentes (38) y (39), obtenemos las fórmulas para calcular $p_N^{(k)}$ y $q_N^{(k)}$:

$$\begin{aligned}
&C_2^{(k-1)}S_N^{(k-1)}=q_N^{(k-1)}+2p_{N-2^{k-1}}^{(k-1)},\\
&p_N^{(k)}=p_N^{(k-1)}+S_N^{(k-1)},\\
&q_N^{(k)}=2p_N^{(k)}+2q_{N-2^{k-1}}^{(k-1)},\quad k=1,\ 2,\ \ldots,\ n,\\
&q_N^{(0)}=F_N,\quad p_N^{(0)}=0.
\end{aligned}\tag{53}$$

Queda por excluir $F_j^{(k)}$ de (35) y (40). Sustituyendo (47) en (35) y (45), (46) en (40), obtenemos las siguientes fórmulas para encontrar Y_j:

$$\mathscr{D}^{(n+1)}S_0^{(n+1)}=q_0^{(n+1)},\qquad Y_0=p_0^{(n+1)}+S_0^{(n+1)},\tag{54}$$

$$C_2^{(n)}S_N^{(n)}=q_N^{(n)}+2Y_0,\qquad Y_N=p_N^{(n)}+S_N^{(n)},\tag{55}$$

$$C^{(k-1)}S_j^{(k-1)}=q_j^{(k-1)}+Y_{j-2^{k-1}}+Y_{j+2^{k-1}},\tag{56}$$

$$Y_j=p_j^{(k-1)}+S_j^{(k-1)},$$

$$j=2^{k-1},\ 3\cdot 2^{k-1},\ \ldots,\ N-2^{k-1},\quad k=n,\ n-1,\ \ldots,\ 1.$$

Así, las fórmulas (48), (52)-(56) describen el método de reducción completa para resolver el tercer problema de contorno (34).

OBSERVACIÓN 1. Si utilizamos las ecuaciones (40′) para hallar Y_0 y Y_N, entonces introduciendo en lugar de $p_0^{(n+1)}$ y $q_0^{(n+1)}$ los vectores $p_N^{(n+1)}$ y $q_N^{(n+1)}$, relacionados con $F_N^{(n+1)}$ por la expresión

$$F_N^{(n+1)} = \mathscr{D}^{(n+1)} p_N^{(n+1)} + q_N^{(n+1)},$$

obtendremos de (38), (42), (44) y (47) las siguientes fórmulas para encontrar $p_N^{(k)}$ y $q_N^{(k)}$:

$$\begin{aligned} &C_2^{(k-1)} S_N^{(k-1)} = q_N^{(k-1)} + 2p_{N-2^{k-1}}^{(k-1)}, \\ &p_N^{(k)} = p_N^{(k-1)} + S_N^{(k-1)}, \quad k = 1, 2, \ldots, n+1, \\ &q_N^{(k)} = 2p_N^{(k)} + 2q_{N-2^{k-1}}^{(k-1)}, \quad k = 1, 2, \ldots, n, \qquad (53') \\ &q_N^{(n+1)} = 4p_N^{(n+1)} + 2q_0^{(n)}, \\ &q_N^{(0)} = F_N, \quad p_N^{(0)} = 0. \end{aligned}$$

Las fórmulas (53′) sustituyen las (53). Ya que en este caso no es necesario calcular el vector $F_0^{(n+1)}$, por consiguiente, tampoco los vectores $p_0^{(n+1)}$ y $q_0^{(n+1)}$, entonces las fórmulas (52) se reemplazan por las siguientes:

$$\begin{aligned} &C_1^{(k-1)} S_0^{(k-1)} = q_0^{(k-1)} + 2p_{2^{k-1}}^{(k-1)}, \quad p_0^{(k)} = p_0^{(k-1)} + S_0^{(k-1)}, \\ &q_0^{(k)} = 2p_0^{(k)} + 2q_{2^{k-1}}^{(k-1)}, \quad k = 1, 2, \ldots, n, \qquad (52') \\ &q_0^{(0)} = F_0, \quad p_0^{(0)} = 0. \end{aligned}$$

De (35) y (40′) obtenemos las fórmulas para hallar Y_0 y Y_N:

$$\mathscr{D}^{(n+1)} S_N^{(n+1)} = q_N^{(n+1)}, \quad Y_N = p_N^{(n+1)} + S_N^{(n+1)}, \qquad (55')$$

$$C_1^{(n)} S_0^{(n)} = q_0^{(n)} + 2Y_N, \quad Y_0 = p_0^{(n)} + S_0^{(n)}. \qquad (54')$$

Las restantes incógnitas se encuentran de acuerdo con (56). De este modo, las fórmulas (48), (52′)-(55′) y (56) también se pueden utilizar para resolver el problema (34).

OBSERVACIÓN 2. Si Y_N está prefijado, es decir, si en lugar de (34) se necesita resolver el problema de contorno

$$\begin{aligned} &(C + 2\alpha E) Y_0 - 2Y_1 = F_0, \quad j = 0, \\ &-Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1, \\ &Y_N = F_N, \quad j = N, \end{aligned}$$

entonces el método de reducción completa en este caso se describe por las fórmulas (48), (52′), (54′) y (56). Si está dado Y_0, es decir, si se resuelve el problema

$$-Y_{j-1} + CY_j - Y_{j+1} = F_j, \quad 1 \leqslant j \leqslant N-1,$$
$$-2Y_{N-1} + (C + 2\beta E)\, Y_N = F_N, \quad j = N, \ Y_0 = F_0,$$

entonces el método se describe mediante las fórmulas (48), (53), (55), y (56).

3.2. FACTORIZACION DE LAS MATRICES. De (39) y (43) se deduce, que $C_1^{(k)}$, $C_2^{(k)}$ y $C^{(k)}$ son los polinomios matriciales de grado 2^k y $\mathscr{B}^{(n+1)}$ de grado 2^{n+1}, con respecto a la matriz C, con coeficiente de mayor grado igual a 1. Factoricemos estas matrices teniendo en cuenta la necesidad de su inversión. Para ello obtendremos una representación explícita de estos polinomios mediante polinomios conocidos y estudiaremos el problema de hallar las raíces de los polinomios indicados.

En el punto 2 del § 2 se mostró, que las $C^{(k)}$ se expresan a través de los polinomios de Chebishev de primer género de la siguiente forma:

$$C^{(k)} = 2T_{2^k}\left(\frac{1}{2}C\right), \quad k = 0,\ 1,\ \ldots\ . \tag{57}$$

Más adelante, de las relaciones (39) hallaremos

$$C_1^{(k)} - C^{(k)} = C^{(k-1)}\,[C_1^{(k-1)} - C^{(k-1)}] = \ldots$$
$$\ldots = \prod_{l=0}^{k-1} C^{(l)}\,[C_1^{(0)} - C^{(0)}] = 2\alpha \prod_{l=0}^{k-1} C^{(l)}. \tag{58}$$

Ya que tiene lugar la igualdad

$$\prod_{l=0}^{k-1} C^{(l)} = \prod_{l=0}^{k-1} 2T_{2^l}\left(\frac{1}{2}C\right) = U_{2^k-1}\left(\frac{1}{2}C\right),$$

donde $U_n(x)$ es el polinomio de Chebishev de segundo género, entonces de (58) obtenemos la siguiente representación para $C_1^{(k)}$:

$$C_1^{(k)} = 2T_{2^k}\left(\frac{1}{2}C\right) + 2\alpha U_{2^k-1}\left(\frac{1}{2}C\right),$$
$$k = 0,\ 1,\ \ldots \tag{59}$$

De modo análogo obtenemos la representación para $C_2^{(k)}$:

$$C_2^{(k)} = 2T_{2^k}\left(\frac{1}{2}C\right) + 2\beta U_{2^k-1}\left(\frac{1}{2}C\right),$$
$$k = 0,\ 1,\ \ldots \tag{60}$$

A continuación, sustituyendo (59) y (60) en (43), tendremos

$$\mathscr{D}^{(n+1)} = 4\left[T_{2^k}\left(\frac{1}{2}C\right)\right]^2 - 4E +$$
$$+ 4(\alpha+\beta)\,T_{2^k}\left(\frac{1}{2}C\right)U_{2^k-1}\left(\frac{1}{2}C\right) +$$
$$+ 4\alpha\beta\left[U_{2^k-1}\left(\frac{1}{2}C\right)\right]^2. \qquad (61)$$

Como tiene lugar la igualdad

$$1 - T_n(x) = U_{n-1}(x)\,(1 - x^2), \qquad (62)$$

entonces de (61) obtenemos

$$\mathscr{D}^{(n+1)} = U_{2^n-1}\left(\frac{1}{2}C\right)\left[(C^2 + 4\alpha\beta E - 4E)\,U_{2^n-1}\left(\frac{1}{2}C\right) + \right.$$
$$\left. + 4(\alpha+\beta)\,T_{2^n}\left(\frac{1}{2}C\right)\right].$$

De esta forma, hemos obtenido la representación para $C^{(h)}$, $C_1^{(h)}$, $C_2^{(h)}$ y $\mathscr{D}^{(n+1)}$ mediante polinomios conocidos. Ya que las raíces de los polinomios de Chebishev de primero y segundo género son conocidas, entonces de (57) y (62) obtenemos

$$C^{(h)} = \prod_{l=1}^{2^k}\left(C - 2\cos\frac{(2l-1)\pi}{2^{k+1}}E\right),$$

$$\mathscr{D}^{(n+1)} = \prod_{l=1}^{2^n-1}\left(C - 2\cos\frac{l\pi}{2^n}E\right)\left[(C^2 + 4\alpha\beta E - 4E)\times\right.$$
$$\left.\times\, U_{2^n-1}\left(\frac{1}{2}C\right) + 4(\alpha+\beta)\,T_{2^n}\left(\frac{1}{2}C\right)\right].$$

Por eso de aquí y de (59), (60) se deduce, que nos quedan por hallar las raíces de los polinomios,

$$P_m(t) = 2T_m\left(\frac{t}{2}\right) + 2\alpha U_{m-1}\left(\frac{t}{2}\right),$$

$$Q_m(t) = 2T_m\left(\frac{t}{2}\right) + 2\beta U_{m-1}\left(\frac{t}{2}\right), \qquad (63)$$

$$m = 2^k,\ k = 0,\ 1,\ \ldots,\ n-1,$$

los cuales corresponden a los polinomios matriciales $C_1^{(k)}$ y $C_2^{(k)}$, y las raíces del polinomio

$$R_{2^n+1}(t)=(t^2+4\alpha\beta-4)\,U_{2^n-1}\left(\frac{t}{2}\right)+\\ +4(\alpha+\beta)\,T_{2^n}\left(\frac{t}{2}\right), \qquad (64)$$

el cual genera al polinomio $\mathscr{D}^{(n+1)}$.

Este problema puede ser resuelto de dos maneras. El primer camino consiste en utilizar uno de los métodos para hallar aproximadamente las raíces de un polinomio y el segundo camino consiste en la reducción de este problema a la búsqueda de todos los valores propios de ciertas matrices tridiagonales. Detengámonos más detalladamente en el segundo procedimiento.

Designemos mediante $S_k(\lambda)$ el siguiente determinante de k-ésimo orden:

$$S_k(\lambda)=\begin{vmatrix} \lambda+2\alpha & 2 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 1 & \lambda & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & \lambda & 1 & \dots & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & & 1 & \lambda & 1 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 1 & \lambda & 1 \\ 0 & 0 & 0 & 0 & & 0 & 0 & 1 & \lambda \end{vmatrix}, \quad k\geqslant 2$$

y pongamos $S_1(\lambda)=\lambda+2\alpha$. De la definición y la estructura de la matriz correspondiente a $S_k(\lambda)$ hallaremos relaciones recurrentes para $S_k(\lambda)$:

$$\begin{aligned} &S_{k+1}(\lambda)=\lambda S_k(\lambda)-S_{k-1}(\lambda), \quad k\geqslant 2, \\ &S_2(\lambda)=\lambda S_1(\lambda)-2, \quad S_1(\lambda)=\lambda+2\alpha. \end{aligned} \qquad (65)$$

Utilizando las relaciones recurrentes para los polinomios de Chebishev (véase el punto 2, § 4, cap. I).

$$T_{n+1}(x)=2xT_n(x)-T_{n-1}(x), \quad T_1(x)=x, \quad T_0(x)=1,$$
$$U_{n+1}(x)=2xU_n(x)-U_{n-1}(x), \quad U_1(x)=2x, \quad U_0(x)=1$$

y las relaciones (65), obtendremos la representación de $S_m(\lambda)$ mediante los polinomios de Chebishev: $S_m(\lambda)=2T_m\left(\frac{\lambda}{2}\right)+2\alpha U_{m-1}\left(\frac{\lambda}{2}\right)$, $m\geqslant 1$. Comparando esta expresión con (63) encontramos, que las raíces del polinomio $P_m(t)$ coinciden con las raíces del determinante $S_m(\lambda)$, el cual depende de λ como de un parámetro.

El problema de encontrar las raíces de $S_m(\lambda)$ es equivalente a la búsqueda de aquellos valores del parámetro λ, para los cuales el sistema de ecuaciones algebraicas

$$\begin{aligned} &y_{i-1}+\lambda y_i+y_{i+1}=0, \quad 1\leqslant i\leqslant m-1, \\ &(\lambda+2\alpha)\,y_0+2y_1=0, \quad i=0, \\ &y_m=0 \end{aligned} \qquad (66)$$

posee solución no nula. Daremos otra forma de escritura para (66). Utilizando la notación para la segunda derivada de diferencias

$$y_{\bar{x}x,\,i}=\frac{1}{h}\,(y_{x,\,i}-y_{\bar{x},\,i})=\frac{1}{h^2}\,(y_{i+1}-2y_i+y_{i-1}),$$

volvamos a escribir (66) en la siguiente forma:

$$\begin{aligned} &y_{\bar{x}x}+\mu y=0,\ 1\leqslant i\leqslant m-1,\\ &\frac{2}{h}\,y_x+\frac{2\alpha}{h^2}\,y+\mu y=0,\ i=0,\ y_m=0, \end{aligned} \qquad (66')$$

donde λ y μ están relacionados por las expresiones $\lambda=\mu h^2-2$. Así, para encontrar las raíces del polinomio $C_i^{(k)}$ es suficiente resolver el problema (66′) para $m=2^k$, $k=0,\ 1,\ \ldots$.

Por analogía con lo expuesto más arriba se puede mostrar, que las raíces del polinomio $Q_m\,(t)$ se encuentran de la solución del problema

$$\begin{aligned} &y_{\bar{x}x}+\mu y=0,\ 1\leqslant i\leqslant m-1,\\ &-\frac{2}{h}\,y_{\bar{x}}+\frac{2\beta}{h^2}\,y+\mu y=0\ \ i=m,\ y_0=0, \end{aligned} \qquad (67)$$

al mismo tiempo la relación $\lambda=\mu h^2-2$ determina estas raíces.

Para encontrar las raíces del polinomio $R_{2^n+1}\,(t)$, definido en (64), es necesario resolver el siguiente problema en valores propios:

$$\begin{aligned} &y_{\bar{x}x}+\mu y=0,\qquad 1\leqslant i\leqslant 2^n-1,\\ &\frac{2}{h}\,y_x+\frac{2\alpha}{h^2}\,y+\mu y=0,\qquad i=0,\\ &-\frac{2}{h}\,y_{\bar{x}}+\frac{2\beta}{h^2}\,y+\mu=0,\qquad i=2^n, \end{aligned} \qquad (68)$$

y hallar las raíces de la igualdad $\lambda=\mu h^2-2$.

Observemos, que para resolver los problemas (66)-(68) se puede utilizar el conocido QR—algoritmo de resolución del problema completo de valores propios.

Capítulo IV
Método de separación de variables

En este capítulo se estudian variantes del método de separación de variables que se aplica para encontrar la solución de las ecuaciones elípticas reticulares más simples en un rectángulo. En el § 1 se expone el algoritmo de la transformación de Fourier discreta rápida de funciones reales y complejas. En el § 2 se examina la variante clásica del método de separación de variables que utiliza el algoritmo de la transformación de Fourier. En el § 3 está construido un método combinado, el cual comprende la reducción incompleta y la separación de variables. Se examina la aplicación de este método a la resolución de problemas de contorno de diferencias para la ecuación de Poisson de segundo y cuarto grados de exactitud.

§ 1. Algoritmo de la transformación de Fourier discreta

1. Planteamiento del problema. Uno de los métodos para buscar las soluciones de los problemas reticulares multidimensionales que admiten separación de variables es el desarrollo de la solución buscada en suma finita de Fourier según las funciones propias de los respectivos operadores reticulares. La efectividad de este método depende esencialmente de la rapidez con que se puedan calcular los coeficientes de Fourier de la función reticular dada y reconstruir la función buscada por los coeficientes de Fourier dados.

Si por ejemplo, sobre la red $\bar{\omega} = \{x_i = ih,\ 0 \leqslant i \leqslant N,\ hN = l\}$, que contiene $N + 1$ nodos, están dadas la función $f(i)$ y el sistema de funciones ortonormalizadas $\mu_k(i)$, $k = 0, 1, \ldots, N$, y los coeficientes de Fourier de la función $f(i)$ se calculan por las fórmulas

$$\varphi_k = \sum_{i=0}^{n} f(i)\,\mu_k(i)\,h, \quad k = 0, 1, \ldots, N, \tag{1}$$

entonces para computar todos los coeficientes φ_k es suficiente $(N+1)(N+2)$ operaciones de multiplicación y $N(N+1)$ operaciones de suma.

En el caso general de un sistema de funciones $\{\mu_k(t)\}$ arbitrario ésta es la cantidad mínima necesaria de operaciones aritméticas. En una serie de casos especiales, cuando el sistema ortonormalizado de funciones posee un tipo especial, el número total de operaciones aritméticas, necesario para el cálculo de las sumas de la forma (1), puede ser reducido sustancialmente. Nosotros examinaremos estos casos y mostraremos algoritmos que permiten calcular todos los coeficientes de Fourier y restablecer la función por los coeficientes de Fourier dados con un gasto de $O(N \ln N)$ operaciones aritméticas.

Pasemos a la descripción de los casos señalados.

PROBLEMA 1. *Desarrollo en senos.* Sea $\bar{\omega} = \{x_j jh, 0 \leqslant j \leqslant N, hN = l\}$ una red uniforme con paso h introducida sobre el segmento $0 \leqslant x \leqslant l$. Designemos mediante $\omega = \{x_j = jh, 1 \leqslant j \leqslant N-1\}$ el conjunto de los nodos interiores de la red $\bar{\omega}$.

Sea $f(j)$ una función real reticular prefijada sobre ω (o $f(t)$ definida sobre $\bar{\omega}$ y al mismo tiempo $f(0) = f(N) = 0$).

En el § 5 del cap. I fue mostrado que la función $f(j)$ puede ser representada en forma del desarrollo

$$f(j) = \frac{2}{N} \sum_{k=1}^{N-1} \varphi_k \operatorname{sen} \frac{k\pi j}{N}, \quad j = 1, 2, \ldots, N-1, \qquad (2)$$

donde los coeficientes φ_k se determinan por la fórmula

$$\varphi_k = \sum_{j=1}^{N-1} f(j) \operatorname{sen} \frac{k\pi j}{N}, \quad k = 1, 2, \ldots, N-1. \qquad (3)$$

Comparando (2) y (3) encontramos, que los problemas de calcular los coeficientes φ_k de una función $f(j)$ dada y de la reconstrucción de esta función por los $\{\varphi_k\}$ dados se reducen al cálculo de $N-1$ sumas del tipo

$$y_k = \sum_{j=1}^{N-1} a_j \operatorname{sen} \frac{k\pi j}{N}, \quad k = 1, 2, \ldots, N-1. \qquad (4)$$

La fórmula (4) describe la regla de transformación de la función reticular a_j, $1 \leqslant j \leqslant N-1$, definida sobre la red

ω, en la función reticular y_j, $1 \leqslant j \leqslant N-1$. La interpretación algebráica de (4) es la siguiente: si designamos mediante $a = (a_1, a_2, \ldots, a_{N-1})$ un vector de dimensión $N-1$, entonces (4) describe la transformación del vector a al pasar de la base natural a la base formada por el sistema de vectores ortogonales

$$z_k = (z_k(1),\ z_k(2),\ \ldots,\ z_k(N-1)),\quad z_k(j) = \\ = \operatorname{sen} \frac{k\pi j}{N}.$$

PROBLEMA 2. *Desarrollo en senos desplazados.* Sea la función reticular $f(j)$, que toma valores reales, definida sobre el conjunto $\omega^+ = \{x_j = jh,\ 1 \leqslant j \leqslant N\}$ (o sobre $\bar{\omega}$ y al mismo tiempo $f(0) = 0$). En el § 5 del cap. I fué mostrado, que tal función $f(j)$ puede ser representada en la forma

$$f(j) = \frac{2}{N} \sum_{k=1}^{N} \varphi_k \operatorname{sen} \frac{(2k-1)\pi j}{2N},\quad j = 1,\ 2,\ \ldots,\ N, \tag{5}$$

donde los coeficientes φ_k se determinan por la fórmula

$$\varphi_k = \sum_{j=1}^{N} \rho_j f(j) \operatorname{sen} \frac{(2k-1)\pi j}{2N},\quad k = 1,\ 2,\ \ldots,\ N, \tag{6}$$

y

$$\rho_j = \begin{cases} 1, & j \neq 0,\ N, \\ 0{,}5, & j = 0,\ N. \end{cases} \tag{7}$$

Si la función $f(j)$ está definida sobre el conjunto $\omega^- = \{x_j = jh,\ 0 \leqslant j \leqslant N-1\}$ (o sobre $\bar{\omega}$ y al mismo tiempo $f(N) = 0$), entonces el desarrollo análogo a (5) y (6) tiene la forma

$$f(N-j) = \frac{2}{N} \sum_{k=1}^{N} \varphi_k \operatorname{sen} \frac{(2k-1)\pi j}{2N},\quad j = 1,\ 2,\ \ldots,\ N, \tag{8}$$

$$\varphi_k = \sum_{j=1}^{N} \rho_{N-j} f(N-j) \operatorname{sen} \frac{(2k-1)\pi j}{2N},\quad k = 1, 2,\ \ldots,\ N, \tag{9}$$

donde la función ρ_j está definida en (7).

De (5), (6), (8) y (9) se deduce, que aquí aparecen problemas del cálculo de sumas del tipo

$$y_k = \sum_{j=1}^{N} a_j \operatorname{sen} \frac{(2k-1)\pi j}{2N}, \quad k = 1, 2, \ldots, N, \tag{10}$$

$$y_j = \sum_{k=1}^{N} a_k \operatorname{sen} \frac{(2k-1)\pi j}{2N}, \quad j = 1, 2, \ldots, N. \tag{10'}$$

PROBLEMA 3. *Desarrollo en cosenos.* Sea $f(j)$ una función real reticular definida sobre la red $\bar{\omega}$. Entonces para la función $f(j)$ tiene lugar el desarrollo

$$f(j) = \frac{2}{N} \sum_{k=0}^{N} \rho_k \varphi_k \cos \frac{k\pi j}{N}, \quad j = 0, 1, \ldots, N, \tag{11}$$

donde

$$\varphi_k = \sum_{j=0}^{N} \rho_j f(j) \cos \frac{k\pi j}{N}, \quad k = 0, 1, \ldots, N, \tag{12}$$

y ρ_j está definido en (7). De las fórmulas (11) y (12) se deduce el problema del cálculo de sumas del tipo

$$y_k = \sum_{j=0}^{N} a_j \cos \frac{k\pi j}{N}, \quad k = 0, 1, \ldots, N. \tag{13}$$

PROBLEMA 4. *Transformación de una función real periódica reticular.* Sea la red uniforme $\Omega = \{x_j = jh,\ j = 0, \pm 1, \pm 2, \ldots, Nh = l\}$ con paso h prefijada sobre el eje $-\infty < x < \infty$. Supongamos que sobre la red Ω está dada una función reticular periódica con período N

$$f(j) = f(j+N), \quad j = 0, \pm 1, \ldots,$$

a cual toma valores reales: en el § 5 del cap. I fue mostrado, que para $0 \leqslant j \leqslant N-1$ la función $f(j)$ es representable en la forma (para N par)

$$f(j) = \frac{2}{N} \left[\sum_{k=0}^{N/2} \rho_k \varphi_k \cos \frac{2k\pi j}{N} + \sum_{k=0}^{N/2-1} \bar{\varphi}_k \operatorname{sen} \frac{2k\pi j}{N}, \right.$$

$$j = 0, 1, \ldots, N-1, \tag{14}$$

donde los coeficientes φ_k y $\bar{\varphi}_k$ se definen por las fórmulas

$$\varphi_k = \sum_{j=0}^{N-1} f(j) \cos \frac{2k\pi j}{N}, \quad k = 0,\ 1,\ \ldots,\ \frac{N}{2}, \tag{15}$$

$$\bar{\varphi}_k = \sum_{j=1}^{N-1} f(j) \operatorname{sen} \frac{2k\pi j}{N}, \quad k = 1,\ 2,\ \ldots,\ \frac{N}{2} - 1, \tag{16}$$

y la función ρ_k es

$$\rho_k = \begin{cases} 1, & k \neq 0,\ N/2, \\ 0{,}5, & k = 0,\ N/2. \end{cases}$$

Las fórmulas (14)-(16) nos conducen al problema de calcular sumas de los tres tipos:

$$\left. \begin{aligned} & y_k = \sum_{j=0}^{N/2} a_j \cos \frac{2k\pi j}{N} + \sum_{j=1}^{N/2-1} \bar{a}_j \operatorname{son} \frac{2k\pi j}{N}, \\ & k = 0,\ 1,\ \ldots,\ N-1, \end{aligned} \right. \tag{17}$$

$$\left. \begin{aligned} y_k &= \sum_{j=0}^{N-1} a_j \cos \frac{2k\pi j}{N}, \quad k = 0,\ 1,\ \ldots,\ N/2, \\ \bar{y}_k &= \sum_{j=1}^{N-1} a_j \operatorname{sen} \frac{2k\pi j}{N}, \quad k = 1,\ 2,\ \ldots,\ N/2 - 1, \end{aligned} \right\} \tag{18}$$

y a su vez los coeficientes a_j en las sumas (18) son los mismos.

PROBLEMA 5. *Transformación de una función compleja periódica reticular.* Supongamos que la función reticular $f(j)$ con período N, definida sobre la red Ω, toma ahora valores complejos. Entonces para $0 \leqslant j \leqslant N-1$ la función $f(j)$ puede ser representada en la forma

$$f(j) = \frac{1}{N} \sum_{k=0}^{N-1} \varphi_k e^{\frac{2k\pi j}{N} i}, \quad j = 0,\ 1,\ \ldots,\ N-1,\ i = \sqrt{-1}, \tag{19}$$

donde los coeficientes complejos φ_k están definidos por la fórmula

$$\varphi_k = \sum_{j=0}^{N-1} f(j)\, e^{-\frac{2k\pi j}{N} i}, \quad k = 0,\ 1,\ \ldots,\ N-1. \tag{20}$$

Notemos, que $\varphi_0 = \varphi_N$ y además,

$$\varphi_{N-k} = \sum_{j=0}^{N-1} f(j) e^{\frac{2k\pi j}{N} i}, \quad k=0,\ 1,\ \ldots,\ N-1.$$

Por eso el cálculo de los coeficientes φ_k y la reconstrucción de la función $f(j)$ se reduce al cálculo de una suma del tipo

$$y_k = \sum_{j=0}^{N-1} a_j e^{\frac{2k\pi j}{N} i}, \quad k=0,\ 1,\ \ldots,\ N-1 \tag{21}$$

con a_j complejos.

Así pues, nos es necesario construir algoritmos para calcular sumas del tipo (4), (10), (13), (17), (18) y (21), que requieran de una cantidad menor que $O(N^2)$ de operaciones aritméticas. Muy sencillamente se construyen algoritmos para el caso, cuando N es una potencia de 2: $N=2^n$, y nosotros nos limitaremos solamente a este caso.

2. **Desarrollo en senos y senos desplazados.** Examinemos detalladamente el algoritmo de cálculo de las sumas (4), suponiendo que $N = 2^n$. En este caso (4) posee la forma

$$y_k = \sum_{j=1}^{2^n-1} a_j^{(0)} \operatorname{sen} \frac{k\pi j}{2^n}, \quad k = 1,\ 2,\ \ldots,\ 2^n-1, \tag{22}$$

donde se ha introducido la notación $a_j^{(0)} = a_j$.

La idea del método consiste en que en la suma (22) los términos con un factor común se agrupan antes de realizarse la multiplicación. En el primer paso del algoritmo se agrupan los términos de las sumas (22) con índices j y $2^n - j$ para $j = 1,\ 2,\ \ldots,\ 2^{n-1} - 1$, y al mismo tiempo se utiliza la igualdad

$$\operatorname{sen} \frac{k\pi (2^n - j)}{2^n} = (-1)^{k-1} \operatorname{sen} \frac{k\pi j}{2^n}. \tag{23}$$

Para esto escribamos (22) en forma de tres sumandos

$$y_k = \sum_{j=1}^{2^{n-1}-1} a_j^{(0)} \operatorname{sen} \frac{k\pi j}{2^n} + \\ + \sum_{j=2^{n-1}+1}^{2^n-1} a_j^{(0)} \operatorname{sen} \frac{k\pi j}{2^n} + a_{2^{n-1}}^{(0)} \operatorname{sen} \frac{k\pi}{2}$$

y realicemos el cambio $j' = 2^n - j$ en la segunda suma. Teniendo en cuenta (23), obtenemos

$$y_k = \sum_{j=1}^{2^{n-1}-1} [a_j^{(0)} + (-1)^{k-1} a_{2^n-j}^{(0)}] \operatorname{sen} \frac{k\pi j}{2^n} + \\ + a_{2^{n-1}}^{(0)} \operatorname{sen} \frac{k\pi}{2}. \qquad (24)$$

Si designamos

$$a_j^{(1)} = a_j^{(0)} - a_{2^n-j}^{(0)},$$

$$a_{2^n-j}^{(1)} = a_j^{(0)} + a_{2^n-j}^{(0)}, \qquad j = 1, 2, \ldots, 2^{n-1} - 1$$

$$a_{2^{n-1}}^{(1)} = a_{2^{n-1}}^{(0)},$$

entonces de (24) tendremos

$$y_{2k-1} = \sum_{j=1}^{2^{n-1}} a_{2^n-j}^{(1)} \operatorname{sen} \frac{(2k-1)\pi j}{2^n}, \quad k = 1, 2, \ldots, 2^{n-1}, \quad (25)$$

$$y_{2k} = \sum_{1=j}^{2^{n-1}-1} a_j^{(1)} \operatorname{sen} \frac{k\pi j}{2^{n-1}}, \quad k = 1, 2, \ldots, 2^{n-1} - 1, \quad (26)$$

De esta manera, como resultado del primer paso tenemos dos sumas del tipo (25) y (26), cada una de las cuales contiene aproximadamente dos veces menos sumandos que la suma inicial (22). Además, las sumas del tipo (26) y la suma inicial poseen una estructura análoga. Por eso a (26) se le puede aplicar el procedimiento de agrupación de los sumandos descrito más arriba.

En el segundo paso, al igual que más arriba, con ayuda de una partición de la suma (26) en tres sumandos y teniendo en cuenta la igualdad (23), donde n se cambia por $n-1$, se agrupan los términos de la suma (26) con índices j y $2^{n-1} - j$ para $j = 1, 2, \ldots, 2^{n-2} - 1$. Como resultado del segundo paso en lugar de (26) obtendremos

$$y_{2(2k-1)} = \sum_{j=1}^{2^{n-2}} a_{2^{n-1}-j}^{(2)} \operatorname{sen} \frac{(2k-1)\pi j}{2^{n-1}}, \quad k = 1, 2, \ldots, 2^{n-2}, \quad (27)$$

$$y_{2^2 k} = \sum_{j=1}^{2^{n-2}-1} a_j^{(2)} \operatorname{sen} \frac{k\pi j}{2^{n-2}}, \quad k = 1, 2, \ldots, 2^{n-2} - 1, \quad (28)$$

donde

$$a_j^{(2)} = a_j^{(1)} - a_{2^{n-1}-j}^{(1)},$$
$$a_{2^{n-1}-j}^{(2)} = a_j^{(1)} + a_{2^{n-1}-j}^{(1)}, \quad j = 1, 2, \ldots, 2^{n-2} - 1,$$
$$a_{2^{n-2}}^{(2)} = a_{2^{n-2}}^{(1)}.$$

De este modo, el problema inicial (22) es equivalente al cálculo de las sumas (25), (27) y (28). La fórmula (28) permite calcular y_k para los k, múltiplos de 4, (27) — para los k, múltiplos de 2, pero no múltiplos de 4 y la fórmula (25) se utiliza para calcular y_k con k impar.

Continuando el proceso de transformación de las sumas que aparecen, obtendremos como resultado del p-ésimo paso

$$y_{2^{s-1}(2k-1)} = \sum_{j=1}^{2^{n-s}} a_{2^{n-s+1}-j}^{(s)} \operatorname{sen} \frac{(2k-1)\pi j}{2^{n-s+1}},$$
$$k = 1, 2, \ldots, 2^{n-s}, \quad s = 1, 2, \ldots, p, \tag{29}$$
$$y_{2^p k} = \sum_{j=1}^{2^{n-p}-1} a_j^{(p)} \operatorname{sen} \frac{k\pi j}{2^{n-p}}, \quad k = 1, 2, \ldots, 2^{n-p} - 1,$$

donde $p = 1, 2, \ldots, n-1$ y los coeficientes $a_j^{(p)}$ se determinan por recurrencia

$$a_j^{(p)} = a_j^{(p-1)} - a_{2^{n-p+1}-j}^{(p-1)},$$
$$a_{2^{n-p+1}-j}^{(p)} = a_j^{(p-1)} + a_{2^{n-p+1}-j}^{(p-1)},$$
$$j = 1, 2, \ldots, 2^{n-p} - 1, \tag{30}$$
$$a_{2^{n-p}}^{(p)} = a_{2^{n-p}}^{(p-1)}, \quad p = 1, 2, \ldots, n-1.$$

Poniendo $p = n-1$ en (29), hallaremos

$$y_{2^{n-1}} = \sum_{j=1}^{1} a_j^{(n-1)} \operatorname{sen} \frac{\pi j}{2} = a_1^{(n-1)},$$
$$y_{2^{s-1}(2k-1)} = \sum_{j=1}^{2^{n-s}} a_{2^{n-s+1}-j}^{(s)} \operatorname{sen} \frac{(2k-1)\pi j}{2^{n-s+1}}, \tag{31}$$
$$k = 1, 2, \ldots, 2^{n-s}$$

para $s = 1, 2, \ldots, n-1$.

Así, el problema inicial (22) ha sido reducido al cálculo del $(n-1)$-ésimo grupo de las sumas (31). La transformación necesaria para esto de los coeficientes $a_j^{(0)}$ se describe mediante las fórmulas (30).

La segunda etapa del algoritmo consiste en la transformación de las sumas (31), las cuales después de sustituir para cada s fijo

$$z_k^{(0)}(1)=y_{2^{s-1}(2k-1)},\quad k=1,\ 2,\ \ldots,\ 2^{n-s},$$
$$b_j^{(0)}(1)=a^{(s)}_{2^{n-s+1}-j},\quad j=1,\ 2,\ \ldots,2^{n-s},$$
$$l=n-s,\qquad s=1,\ 2,\ \ldots,\ n-1.$$

se escriben en la siguiente forma:

$$z_k^{(0)}(1)=\sum_{j=1}^{2^l} b_j^{(0)}(1)\operatorname{sen}\frac{(2k-1)\pi j}{2^{l+1}},\quad k=1,\ 2,\ \ldots,\ 2^l, \tag{32}$$

donde $l=1,\ 2,\ \ldots,\ n-1$. Aquí los coeficientes $b_j^{(0)}(1)$ y las funciones $z_k^{(0)}(1)$ dependen del índice l, pero como nosotros expondremos el método de cálculo de la suma (32) para l fijo, entonces este índice ha sido omitido en todas partes.

Ocupémonos de la transformación de la suma (32). Representémosla en forma de dos sumandos, separando los términos con índices j pares e impares:

$$z_k^{(0)}(1)=\sum_{j=1}^{2^{l-1}} b_{2j}^{(0)}(1)\operatorname{sen}\frac{(2k-1)\pi j}{2^l}+$$
$$+\sum_{j=1}^{2^{l-1}} b_{2j-1}^{(0)}(1)\operatorname{sen}\frac{(2k-1)\pi(2j-1)}{2^{l+1}}. \tag{33}$$

Utilizando la igualdad

$$\operatorname{sen}\frac{(2k-1)(2j-1)\pi}{2^{l+1}}+\operatorname{sen}\frac{(2k-1)2j\pi}{2^{l+1}}=$$
$$=2\cos\frac{(2k-1)\pi}{2^{l+1}}\operatorname{sen}\frac{(2k-1)(2j-1)\pi}{2^{l+1}},$$

escribamos el segundo sumando en forma de dos sumas

$$\sum_{j=1}^{2^{l-1}} b_{2j-1}^{(0)}(1) \operatorname{sen} \frac{\pi(2k-1)(2j-1)}{2^{l+1}} =$$

$$= \frac{1}{2\cos\frac{(2k-1)\pi}{2^{l+1}}}\left[\sum_{j=1}^{2^{l-1}} b_{2j-1}^{(0)}(1) \operatorname{sen} \frac{(2k-1)\pi j}{2^{l}} + \right.$$

$$\left. + \sum_{j=1}^{2^{l-1}} b_{2j-1}^{(0)}(1) \operatorname{sen} \frac{(2k-1)\pi(j-1)}{2^{l}}\right] =$$

$$= \frac{1}{2\cos\frac{(2k-1)\pi}{2^{l+1}}}\left(b_{2^l-1}^{(0)}(1) \operatorname{sen} \frac{(2k-1)\pi}{2} + \right.$$

$$\left. + \sum_{i=1}^{2^{l-1}-1} \left(b_{2j+1}^{(0)}(1) + b_{2j-1}^{(0)}(1)\right) \operatorname{sen} \frac{(2k-1\pi j}{2^{l}}\right). \quad (34)$$

Aclaremos, que en la segunda suma, situada en los corchetes, fué hecho el cambio del índice $j = j' + 1$.

Designemos

$$b_j^{(1)}(1) = b_{2j-1}^{(0)}(1) + b_{2j+1}^{(0)}(1), \quad j = 1, 2, \ldots, 2^{l-1} - 1,$$

$$b_{2^l-1}^{(1)}(1) = b_{2^l-1}^{(0)}(1),$$

$$b_j^{(1)}(2) = b_{2j}^{(0)}(1), \qquad j = 1, 2, \ldots 2^{l-1}$$

y sustituyamos (34) en (33). Obtenemos la expresión

$$z_k^{(0)}(1) = \sum_{j=1}^{2^{l-1}} b_j^{(1)}(2) \operatorname{sen} \frac{(2k-1)\pi j}{2^{l}} +$$

$$+ \frac{1}{2\cos\frac{(2k-1)\pi}{2^{l+1}}} \sum_{j=1}^{2^{l-1}} b_j^{(1)}(1) \operatorname{sen} \frac{(2k-1)\pi j}{2^{l}},$$

válida para $k = 1, 2 \ldots, 2^l$. Poniendo aquí en lugar de k el indice $2^l - k + 1$, obtendremos

$$z_{2^l-k+1}^{(0)}(1) = -\sum_{j=1}^{2^{l-1}} b_j^{(1)}(2) \operatorname{sen} \frac{(2k-1)\pi j}{2^{l}} +$$

$$+ \frac{1}{2\cos\frac{(2k-1)\pi}{2^{l+1}}} \sum_{j=1}^{2^{l-1}} b_j^{(1)}(1) \operatorname{sen} \frac{(2k-1)\pi j}{2^{l}}.$$

Por consiguiente, si designamos

$$z_k^{(1)}(s) = \sum_{j=1}^{2^{l-1}} b_j^{(1)}(s) \operatorname{sen} \frac{(2k-1)\pi j}{2^l},$$

$$k = 1, 2, \ldots, 2^{l-1}, \quad s = 1, 2,$$

entonces la suma inicial $z_k^{(0)}(1)$ puede ser calculada por las fórmulas

$$z_k^{(0)}(1) = z_k^{(1)}(2) + \frac{1}{2\cos\frac{(2k-1)\pi}{2^{l+1}}} z_k^{(1)}(1),$$

$$z_{2^l-k+1}^{(0)}(1) = -z_k^{(1)}(2) + \frac{1}{2\cos\frac{(2k-1)\pi}{2^{l+1}}} z_k^{(1)}(1),$$

$$k = 1, 2, \ldots, 2^{l-1}.$$

De esta forma, el primer paso ha proporcionado la aparición de las sumas $z_k^{(1)}(1)$ y $z_k^{(1)}(2)$, cada una de las cuales contiene dos veces menos sumandos, que la suma inicial $z_k^{(0)}(1)$, pero tiene la misma estructura que $z_k^{(0)}(1)$. En virtud de esto el proceso de transformación de la suma inicial descrito más arriba puede ser aplicado por separado a las sumas $z_k^{(1)}(1)$ y $z_k^{(1)}(2)$. Como resultado aparecerán las sumas $z_k^{(2)}(s)$, $s = 1, 2, 3, 4$, que conservan la estructura de la suma inicial. Continuando el proceso de transformaciones, en el m-ésimo paso obtendremos las sumas

$$z_k^{(m)}(s) = \sum_{j=1}^{2^{l-m}} b_j^{(m)}(s) \operatorname{sen} \frac{(2k-1)\pi j}{2^{l-m+1}} \tag{35}$$

$$k = 1, 2, \ldots, 2^{l-m}, \quad s = 1, 2, \ldots, 2^m$$

para cada $m = 0, 1, \ldots, l$, donde los coeficientes $b_j^{(m)}(s)$ se determinan por recurrencia para $s = 1, 2, \ldots, 2^{m-1}$ según las fórmulas

$$b_j^{(m)}(2s-1) = b_{2j-1}^{(m-1)}(s) - b_{2j+1}^{(m-1)}(s),$$
$$j = 1, 2, \ldots, 2^{l-m} - 1, \quad m = 1, 2, \ldots, l-1,$$
$$b_{2^{l-m}}^{(m)}(2s-1) = b_{2^{l-m+1}-1}^{(m-1)}(s), \quad m = 1, 2, \ldots, l, \tag{36}$$
$$b_j^{(m)}(2s) = b_{2j}^{(m-1)}(s), \quad j = 1, 2, \ldots, 2^{l-m}, \quad m = 1, 2, \ldots, l.$$

A su vez las sumas del m-ésimo paso están relacionadas con las sumas obtenidas en el $(m-1)$-ésimo paso, por las siguientes fórmulas:

$$z_k^{(m-1)}(s) = z_k^{(m)}(2s) + \frac{1}{2\cos\frac{\pi(2k-1)}{2^{l-m+2}}} z_k^{(m)}(2s-1),$$

$$z_{2^{l-m+1}-k+1}^{(m-1)}(s) = -z_k^{(m)}(2s) + \frac{1}{2\cos\frac{\pi(2k-1)}{2^{l-m+2}}} z_k^{(m)}(2s-1), \quad (37)$$

$k = 1, 2, \ldots, 2^{(l)}, \quad s = 1, 2, \ldots, 2^{m-1},$

$m = 1, 2, \ldots, l.$

Poniendo $m = l$ en (35), obtendremos

$$z_1^{(l)}(s) = b_1^{(l)}(s), \quad s = 1, 2, \ldots, 2^l. \quad (38)$$

Así pues, las sumas $z_k^{(0)}(1)$ se calculan de la siguiente forma. Partiendo de los coeficientes dados $b_j^{(0)}(1)$, $1 \leqslant j \leqslant 2^l$, por las fórmulas (36) se computan en total los coeficiente- $b_1^{(l)}(s)$, $1 \leqslant s \leqslant 2^l$. En virtud de (38) ellos se utilizan después en calidad de datos iniciales para las relaciones recu. rrentes (37). Poniendo en (37) sucesivamente $m = l, l-1, \ldots$ s . . ., 1, obtendremos como resultado $z_k^{(0)}(1)$ y, por lo tanto, $y_{2^{s-1}\,2k-1}$.

De esta forma, el algoritmo del cálculo de las sumas (22) se describe por las fórmulas (30), (36), (38).

OBSERVACIÓN. En las relaciones recurrentes (37) se puede evitar la división por $2\cos\frac{(2k-1)\pi}{2^{l-m+2}}$ por medio de la sustitución

$$z_k^{(m)}(s) = \operatorname{sen}\frac{(2k-1)\pi}{2^{l-m+1}} w_k^{(m)}(s).$$

En este caso las fórmulas para calcular $w^{(m)}$ toman la forma

$$w_k^{(m-1)}(s) = 2\cos\frac{\pi(2k-1)}{2^{l-m+2}} w_k^{(m)}(2s) + w_k^{(m)}(2s-1),$$

$$w_{2^{l-m+1}-k+1}^{(m-1)}(s) = -2\cos\frac{\pi(2k-1)}{2^{l-m+2}} w_k^{(m)}(2s) + w_k^{(m)}(2s-1), \quad (39)$$

$k = 1, 2, \ldots, 2^{l-m}, \quad s = 1, 2, \ldots, 2^{m-1}, \quad m = l, l-1, \ldots, 1,$

al mismo tiempo $w_1^{(l)}(s)=b_1^{(l)}(s)$, $s=1, 2, \ldots, 2^l$ y

$$z_k^{(0)}(1)=\operatorname{sen}\frac{(2k-1)\pi}{2^{l+1}}\,w_k^{(0)}(1), \quad k=1, 2, \ldots, 2^l. \qquad (40)$$

Contemos ahora el número de operaciones aritméticas que es necesario ejecutar para la realización del algoritmo (30), (36)-(38). Supondremos, que los valores de las funciones trigonométricas están calculados con anticipación.

Un cálculo elemental da

1) en la realización de (30) se exige

$$Q_1=\sum_{p=1}^{n-1}2(2^{n-p}-1)=2\cdot 2^n-2(n+1)$$

operaciones de suma y resta;

2) para l fijo en la realización de (36) se exige

$$\bar{q}_l=\sum_{m=1}^{l-1}(2^{l-m}-1)\cdot 2^{m-1}=(l-2)\,2^{l-1}+1$$

operaciones de suma, y en la realización de (37) se exige

$$\bar{\bar{q}}_l=\sum_{m=1}^{l}2\cdot 2^{l-m}\cdot 2^{m-1}=2l\cdot 2^{l-1}$$

operaciones de suma y

$$q_l^*=\sum_{m=1}^{l}2^{l-m}\cdot 2^{m-1}=l\cdot 2^{l-1} \qquad (41)$$

operaciones de multiplicación. En total las fórmulas (36) y (37) exigen para l fijo

$$q_l=\bar{q}_l+\bar{\bar{q}}_l=(3l-2)\cdot 2^{l-1}+1 \qquad (42)$$

operaciones de suma y q_l^* multiplicaciones. Para todos los $l=1, 2, \ldots, n-1$ los gastos constituyen

$$Q_2=\sum_{l=1}^{n-1}q_l=\sum_{l=1}^{n-1}[(3l-2)\cdot 2^{l-1}+1]=\frac{3}{2}\,n\,2^n-4\cdot 2^n+n+4$$

sumas y

$$Q_3=\sum_{l=1}^{n-1}q_l^*=\sum_{l=1}^{n-1}l2^{l-1}=\frac{n}{2}\,2^n-2^n+1$$

multiplicaciones.

De esta manera, el algoritmo (30), (36)-(38) se caracteriza por las siguientes estimaciones del número de operaciones aritméticas: $Q_+ = Q_1 + Q_2 = (3n/2 - 2)\, 2^n - n + 2$ sumas y $Q_* = (n/2 - 1)\, 2^n + 1$ multiplicaciones. Si no hacemos diferencia entre las operaciones de suma y producto, entonces el número total de operaciones es

$$Q = Q_1 + Q_2 + Q_3 = (2 \log_2 N - 3)\, N - \log_2 N + 3,$$
$$N = 2^n.$$

Para comparar citemos la estimación del número de operaciones, que es necesario efectuar, para calcular todas las sumas (22) por sumación directa. Tendremos $(2^n - 1)^2$ operaciones de multiplicación y $(2^n - 2)(2^n - 1)$ operaciones de suma, y por todo $\tilde{Q} = (N - 1)(2N - 3)$. Por ejemplo, para $N = 128$ $(n = 7)$ obtenemos $Q = 1404$ operaciones (de ellas 321 operaciones de multiplicación) para el algoritmo construido y $Q = 32\,131$ operaciones (de ellas 15 873 operaciones de multiplicación) para el algoritmo de sumación directa.

Indiquemos, que la utilización de (39) y (40) en el algoritmo en lugar de (37) y (38) conduce a las siguientes estimaciones del número de operaciones:
$Q_+ = (\frac{3}{2} n - 2)\, 2^n - n + 2$ sumas y $Q_* = \frac{n}{2}\, 2 \;\; - 1$ multiplicaciones, y en total $Q = (2 \log_2 N - 2)\, N - \log_2 N + 1$, $N = 2^n$, lo cual es algo más que en el algoritmo (30), (36)-(38).

Así, está resuelto el problema 1 planteado más arriba. Examinemos ahora el problema 2 sobre el desarrollo en senos desplazados. Suponiendo que $N = 2^n$, escribamos la suma que figura en el problema 2, en la siguiente forma:

$$y_k = \sum_{j=1}^{2^n} a_j \operatorname{sen} \frac{(2k-1)\pi j}{2^{n+1}}, \quad k = 1, 2, \ldots, 2^n. \tag{43}$$

Comparando (43) con (32), encontramos, que si en (32) ponemos $l = n$, entonces el cálculo de las sumas (43) por los senos desplazados es la segunda etapa del dicho algoritmo para calcular las sumas (22). Por lo tanto, si designamos

$$z_k^{(0)}(1) = y_k, \quad k = 1, 2, \ldots, 2^n,$$
$$b_j^{(0)}(1) = a_j, \quad j = 1, 2, \ldots, 2^n,$$

entonces las fórmulas (36)-(38) para $l = n$ describen el algoritmo del cálculo de las sumas (43). Poniendo $l = n$ en las fórmulas (41) y (42), obtendremos las siguientes estimaciones para el algoritmo construido.

$Q_+ = q_n = (\frac{3}{2}n - 1)\, 2^n + 1$ operaciones de suma y $Q_* = q_n^* = \frac{n}{2}\, 2^n$ operaciones de multiplicación, y en total $Q = (2 \log_2 N - 1)\, N + 1$, $N = 2^n$. De esta forma, las sumas (43) se calculan aproximadamente con los mismos gastos de operaciones aritméticas que las sumas (22).

Recordemos, que las sumas (43) se utilizan para calcular los coeficientes de Fourier de la función reticular a_i, prefijada para $i = 1, 2, \ldots, N$. Para reconstruir la función por sus coeficientes de Fourier dados es necesario calcular las sumas

$$y_j = \sum_{k=1}^{2^n} a_k \operatorname{sen} \frac{(2k-1)\pi j}{2^{n+1}}, \quad j = 1, 2, \ldots, 2^n. \qquad (43')$$

Utilizando para $j \neq 2^n$ las relaciones

$$\operatorname{sen} \frac{(2k-1)\pi j}{2^{n+1}} = \frac{1}{2 \cos \frac{\pi j}{2^{n+1}}} \left[\operatorname{sen} \frac{(k-1)\pi j}{2^n} + \operatorname{sen} \frac{k\pi j}{2^n} \right],$$

obtenemos

$$y_j = \frac{1}{2 \cos \frac{\pi j}{2^{n+1}}} \left[\sum_{k=1}^{2^n} a_k \operatorname{sen} \frac{(k-1)\pi j}{2^n} + \sum_{k=1}^{2^n} a_k \operatorname{sen} \frac{k\pi j}{2^n} \right] =$$

$$= \frac{1}{2 \cos \frac{\pi j}{2^{n+1}}} \sum_{k=1}^{2^n-1} a_k^{(0)} \operatorname{sen} \frac{k\pi j}{2^n}, \quad j = 1, 2, \ldots, 2^{n-1},$$

donde los $a_k^{(0)}$ se calculan por la fórmula $a_k^{(0)} = a_k + a_{k+1}$, $k = 1, 2, \ldots, 2^n - 1$. Comparando la suma obtenida con (22), encontramos, que el problema se redujo al problema 1 resuelto anteriormente.

Para el cálculo de y_{2^n} obtenemos la fórmula

$$y_{2^n} = \sum_{k=1}^{2^n} a_k (-1)^{k-1} = \sum_{k=1}^{2^{n-1}} (a_{2k-1} - a_{2k}).$$

Aquí la sumación se efectúa directamente.

Para el número de operaciones del algoritmo expuesto es válida la estimación $Q = 2N \log_2 N - \log_2 N$.

3. Desarrollo en cosenos. Examinemos ahora un algoritmo de resolución del problema 3 que consiste en el cálculo de las sumas (13), para $N = 2^n$. Tenemos

$$y_k = \sum_{j=0}^{2^n} a_j^{(0)} \cos \frac{k\pi j}{2^n}, \quad k = 0, 1, \ldots, 2^n, \tag{44}$$

donde hemos introducido la notación $a_j^{(0)} = a_j$.

El principio de construcción del algoritmo es exactamente el mismo que para el desarrollo en senos y consiste de dos etapas. En la primera se agrupan primeramente los sumandos de las sumas con índices j y $2^n - j$ para $j = 0, 1, \ldots$ $\ldots, 2^{n-1} - 1$, luego aquellos con índices j y $2^{n-1} - j$ para $j = 0, 1, \ldots, 2^{n-2} - 1$ y así sucesivamente.

Como resultado del p-ésimo paso tendremos

$$y_{2^{s-1}(2k-1)} = \sum_{j=0}^{2^{n-s-1}} a^{(s)}_{2^{n-s+1}-j} \cos \frac{(2k-1)\pi j}{2^{n-s+1}},$$

$$k\;\; 1, 2, \ldots, 2^{n-s}, \quad s = 1, 2, \ldots, p, \tag{45}$$

$$y_{2^p k} = \sum_{j=0}^{2^{n-p}} a_j^{(p)} \cos \frac{k\pi j}{2^{n-p}}, \quad k = 0, 1, \ldots, 2^{n-p}.$$

Estas fórmulas son válidas para $p = 1, 2, \ldots, n$. Los coeficientes $a_j^{(p)}$ se determinan por recurrencia

$$a_j^{(p)} = a_j^{(p-1)} + a^{(p-1)}_{2^{n-p+1}-j},$$

$$a^{(p)}_{2^{n-p+1}-j} = a_j^{(p-1)} - a^{(p-1)}_{2^{n-p+1}-j}, \; j = 0, 1, \ldots, 2^{n-p} - 1, \tag{46}$$

$$a^{(p)}_{2^{n-p}} = a^{(p-1)}_{2^{n-p}}, \qquad p = 1, 2, \ldots, n.$$

Poniendo $s = p = n$ en (45), hallaremos

$$y_0 = a_0^{(n)} + a_1^{(n)}, \quad y_{2^n} = a_0^{(n)} - a_1^{(n)}, \quad y_{2^{n-1}} = a_2^{(n)}, \tag{47}$$

y los restantes y_k se encuentran por las fórmulas

$$y_{2^{s-1}(2k-1)} = \sum_{j=0}^{2^{n-s}-1} a^{(s)}_{2^{n-s+1}-j} \cos \frac{(2k-1)\pi j}{2^{n-s+1}},$$

$$k = 1, 2, \ldots, 2^{n-s}, \quad s = 1, 2, \ldots, n-1.$$

Los cambios para cada s fijo

$$z_k^{(0)}(1) = y_{2^{s-1}(2k-1)}, \qquad k = 1,\ 2,\ \dots,\ 2^{n-s},$$

$$b_j^{(0)}(1) = a^{(s)}_{2^{n-s+1}-j}, \qquad j = 0,\ 1,\ \dots,\ 2^{n-s}-1,$$

$$l = n - s, \qquad s = 1,\ 2,\ \dots,\ n-1$$

nos conducen al cálculo de las sumas siguientes:

$$z_k^{(0)}(1) = \sum_{j=0}^{2^l-1} b_j^{(0)}(1) \cos \frac{(2k-1)\pi j}{2^{l+1}}, \quad k = 1,\ 2,\ \dots,\ 2^l,$$

$$l = 1,\ 2,\ \dots,\ n-1 \qquad (48)$$

La segunda etapa del algoritmo consiste en el cálculo de las sumas (48). Como antes, separando sucesivamente los sumandos con índices j pares e impares, tendremos las siguientes relaciones recurrentes:

$$z_k^{(m-1)}(s) = z_k^{(m)}(2s) + \frac{1}{2\cos\frac{(2k-1)\pi}{2^{l-m+2}}} z_k^{(m)}(2s-1),$$

$$z^{(m)}_{2^{l-m+1}-k+1}(s) = z_k^{(m)}(2s) - \frac{1}{2\cos\frac{(2k-1)\pi}{2^{l-m+2}}} z_k^{(m)}(2s-1), \qquad (49)$$

$k = 1,\ 2,\ \dots,\ 2^{l-m}$, $s = 1,\ 2,\ \dots,\ 2^{m-1}$, $m = 1,\ 2,\ \dots,\ l$

para calcular

$$z_k^{(m)}(s) = \sum_{j=0}^{2^{l-m}-1} b_j^{(m)}(s) \cos \frac{(2k-1)\pi j}{2^{l-m+1}},$$

$$k = 1,\ 2,\ \dots,\ 2^{l-m},\ x = 1,\ 2,\ \dots,\ 2^m \qquad (50)$$

con $m = 0,\ 1,\ \dots,\ l$. Los coeficientes $b_j^{(m)}(s)$ también se determinan por recurrencia para $s = 1,\ 2,\ \dots,\ 2^{m-1}$, comenzando por $b_j^{(0)}(1)$, según las fórmulas

$$b_j^{(m)}(2s-1) = b_{2j-1}^{(m-1)}(s) + b_{2j+1}^{(m-1)}(s),$$

$$j = 1,\ 2,\ \dots,\ 2^{l-m}-1, \qquad m = 1,\ 2,\ \dots,\ l-1,$$

$$b_0^{(m)}(2s-1) = b_1^{(m-1)}(s), \quad m = 1,\ 2,\ \dots,\ l, \qquad (51)$$

$$b_j^{(m)}(2s) = b_{2j}^{(m-1)}(s),$$

$$j = 0,\ 1,\ \dots,\ 2^{l-m}-1,\ m = 1,\ 2,\ \dots,\ l.$$

Suponiendo $m = l$ en (50), hallaremos los datos iniciales para las relaciones (49)

$$z_1^{(l)}(s) = b_0^{(l)}(s), \quad s = 1, 2, \ldots, 2^l. \tag{52}$$

Así pues, el algoritmo del cálculo de las sumas (44), se describe por las fórmulas (46), (47), (49), (51) y (52).

Un cálculo elemental del número de operaciones aritméticas para el algoritmo construido da: $Q_+ = (3/2n - 2)\,2^n + n + 2$ operaciones de suma y $Q_* = (n/2 - 1)\,2^n + 1$ operaciones de multiplicación, y en total

$$Q = Q_+ + Q_* = (2\log_2 N - 3)\,N + \log_2 N + 3,$$
$$N = 2^n.$$

Notemos que, como en el algoritmo anterior, aquí en las relaciones (49) es posible la sustitución

$$z_k^{(m)}(s) = \operatorname{sen}\frac{(2k-1)\pi}{2^{l-m+1}}\,w_k^{(m)}(s);$$

en este caso de (52) se deduce que $w_1^{(l)}(s) = b_0^{(l)}(s)$, $s = 1, 2, \ldots, 2^l$.

Las fórmulas recurrentes para $w_k^{(m)}(s)$ poseen la forma

$$w_k^{(m-1)}(s) = 2\cos\frac{(2k-1)\pi}{2^{l-m+2}}\,w_k^{(m)}(2s) + w_k^{(m)}(2s-1),$$

$$w_{2^{l-m+1}-k+1}^{(m-1)}(s) = 2\cos\frac{(2k-1)\pi}{2^{l-m+2}}\,w_k^{(m)}(s) - w_k^{(m)}(2s-1),$$

$$k = 1, 2, \ldots, 2^{l-m}, \quad s = 1, 2, \ldots, 2^{m-1}, \quad m = 1, 2, \ldots, l.$$

4. Transformación de una función real periódica reticular. El problema 4 sobre la transformación de una función real periódica reticular consiste en la reconstrucción de una función según las fórmulas (17) para coeficientes de Fourier a_j y $\bar{a}_j$ dados y en encontrar los coeficientes para una función dada por las fórmulas (18).

Sean dados los coeficientes de Fourier y sea $N = 2^n$. Entonces es necesario calcular las sumas

$$y_k = \sum_{j=0}^{2^{n-1}}{}' a_j^{(0)} \cos\frac{2k\pi j}{2^n} + \sum_{j=1}^{2^{n-1}-1} \bar{a}_j^{(0)} \operatorname{sen}\frac{2k\pi j}{2^n},$$
$$k = 0, 1, \ldots, 2^n - 1. \tag{53}$$

Construyamos el algoritmo correspondiente. Para esto cambiemos el índice k por $2^n - k$ en (53). Teniendo en

cuenta las igualdades

$$\cos \frac{2(2^n-k)\pi j}{2^n} = \cos \frac{2k\pi j}{2^n},$$

$$\operatorname{sen} \frac{2(2^n-k)\pi j}{2^n} = -\operatorname{sen} \frac{2k\pi j}{2^n},$$

obtenemos, que y_k se puede calcular por las fórmulas

$$\begin{aligned} &y_k = \bar{y}_k + \bar{\bar{y}}_k, \\ &\bar{y}_{2^n-k} = y_k - \bar{\bar{y}}_k, \quad k = 1, 2, \ldots, 2^{n-1}-1, \\ &y_0 = \bar{y}_0, \quad y_{2^{n-1}} = \bar{y}_{2^{n-1}}, \end{aligned} \tag{54}$$

donde

$$\bar{y}_k = \sum_{j=0}^{2^{n-1}} a_j^{(0)} \cos \frac{k\pi j}{2^{n-1}}, \quad k = 0, 1, \ldots, 2^{n-1}, \tag{55}$$

$$\bar{\bar{y}}_k = \sum_{j=1}^{2^{n-1}-1} \bar{a}_j^{(0)} \operatorname{sen} \frac{k\pi j}{2^{n-1}}, \quad k = 1, 2, \ldots, 2^{n-1}-1. \tag{56}$$

Así, el cálculo de las sumas (53) se reduce al cálculo de las sumas (55) y (56) y a la subsiguiente utilización de las fórmulas (54).

Comparando las fórmulas (55) y (56) con las fórmulas (44) y (22), encontramos que las sumas (55) y (56) se pueden calcular por los algoritmos de los puntos 2 y 3, cambiando en ellos n por $n-1$.

Contemos ahora el número de operaciones aritméticas, necesarias para calcular las sumas (53) mediante el procedimiento indicado. De las estimaciones del número de operaciones, halladas para el algoritmo del punto 2, obtenemos, que las sumas (56) se calculan con un gasto de $Q_+ = (3n/4 - 7/4)\, 2^n - n + 3$ operaciones de suma y $Q_* = (n/4 - 3/4)\, 2^n + 1$ operaciones de multiplicación.

Las estimaciones del algoritmo del punto 3 dan los siguientes valores para las sumas (55): $Q_+ = (3n/4 - 7/4)\, 2^n + n + 1$ operaciones de sumas y $Q_* = (n/4 - 3/4)\, 2^n + 1$ operaciones de multiplicación. Añadiendo aquí las $Q_+ = 2^n - 2$ operaciones de suma gastadas en la realización de (54), obtendremos para el algoritmo construido $Q_+ = (3n/2 - 5/2)\, 2^n + 2$ operaciones de suma $Q_* = (n/2 - 3/2)\, 2^n + 2$ operaciones de multiplicación, y en total $Q = (2 \log_2 N - 4) N + 4$, $N = 2^n$.

Regresemos ahora al cálculo de los coeficientes de Fourier de una función real periódica reticular. El problema consiste en hallar las sumas

$$y_k = \sum_{j=0}^{2^n-1} a_j^{(0)} \cos\frac{2k\pi j}{2^n}, \quad k=0, 1, \ldots, 2^{n-1}, \qquad (57)$$

$$\bar{y}_k = \sum_{j=1}^{2^n-1} a_j^{(0)} \operatorname{sen}\frac{2k\pi j}{2^n}, \quad k=1, 2, \ldots, 2^{n-1}-1, \quad (58)$$

donde $a_j^{(0)}$ es una función prefijada.

El algoritmo de cálculo (57) y (58) es cercano a los algoritmos de los puntos 2 y 3, pero se diferencia por algunos detalles. Aquí en la primera etapa se agrupan primeramente los términos de las sumas (57) y (58) con índices j y $2^{n-1}+j$ para $j=0, 1, \ldots, 2^{n-1}-1$, después con índices j y $2^{n-2}+j$ para $j=0, 1, \ldots, 2^{n-2}-1$ y así sucesivamente. Mostremos con más detalle el primer paso del proceso de agrupación sucesiva de los sumandos en el ejemplo de la suma (57). La transformación de la suma (58) se realiza análogamente.

Así, representemos (57) en la siguiente forma:

$$y_k = \sum_{j=0}^{2^{n-1}-1} a_j^{(0)} \cos\frac{2k\pi j}{2^n} + \sum_{j=2^{n-1}}^{2^n-1} a_j^{(0)} \cos\frac{2k\pi j}{2^n}$$

y efectuemos en la segunda suma una sustitución, poniendo $j = 2^{n-1}+j'$. Esto da

$$y_k = \sum_{j=0}^{2^{n-1}-1} [a_j^{(0)} + (-1)^k a_{2^{n-1}+j}^{(0)}] \cos\frac{2k\pi j}{2^n},$$
$$k=0, 1, \ldots, 2^{n-1}.$$

Designando

$$a_j^{(1)} = a_j^{(0)} + a_{2^{n-1}+j}^{(0)},$$

$$a_{2^{n-1}+j}^{(1)} = a_j^{(0)} - a_{2^{n-1}+j}^{(0)}, \quad j=0, 1, \ldots, 2^{n-1}-1, \quad (59)$$

obtendremos en lugar de (58) las siguientes sumas:

$$y_{2k-1}=\sum_{j=0}^{2^{n-1}-1} a^{(1)}_{2^{n-1}+j}\cos\frac{(2k-1)\pi j}{2^{n-1}},\quad k=1,\ 2,\ \ldots,\ 2^{n-2},$$
$$y_{2k}=\sum_{j=0}^{2^{n-1}-1} a^{(1)}_{j}\cos\frac{2k\pi j}{2^{n-1}},\quad k=0,\ 1,\ \ldots,\ 2^{n-2}. \qquad (60)$$

Análogamente en lugar de (58) obtenemos las sumas

$$\bar y_{2k-1}=\sum_{j=1}^{2^{n-1}-1} a^{(1)}_{2^{n-1}+j}\operatorname{sen}\frac{(2k-1)\pi j}{2^{n-1}},\quad k=1,\ 2,\ \ldots,\ 2^{n-2},$$
$$\bar y_{2k}=\sum_{j=1}^{2^{n-1}-1} a^{(1)}_{j}\operatorname{sen}\frac{2k\pi j}{2^{n-1}},\quad k=1,\ 2,\ \ldots,\ 2^{n-2}-1, \qquad (61)$$

donde los $a_j^{(1)}$ están definidos en (59). Con esto está terminado el primer paso. En el segundo paso por el procedimiento descrito se transforman las sumas (60) y (61). Como resultado del p-ésimo paso tendremos

$$y_{2^{s-1}(2k-1)}=\sum_{j=0}^{2^{n-s}-1} a^{(s)}_{2^{n-s}+j}\cos\frac{(2k-1)\pi j}{2^{n-s}},$$
$$k=1,\ 2,\ \ldots,\ 2^{n-s-1},\quad s=1,\ 2,\ \ldots,\ p, \qquad (62)$$
$$y_{2^p k}=\sum_{j=0}^{2^{n-p}-1} a^{(p)}_{j}\cos\frac{2k\pi j}{2^{n-p}},\quad k=0,\ 1,\ \ldots,\ 2^{n-p-1},$$

donde $p=1,\ 2,\ \ldots,\ n-1$ y

$$\bar y_{2^{s-1}(2k-1)}=\sum_{j=1}^{2^{n-s}-1} a^{(s)}_{2^{n-s}+j}\operatorname{sen}\frac{(2k-1)\pi j}{2^{n-s}},$$
$$k=1,\ 2,\ \ldots,\ 2^{n-s-1},\quad s=1,\ 2,\ \ldots,\ p, \qquad (63)$$
$$\bar y_{2^p k}=\sum_{j=1}^{2^{n-p}-1} a^{(p)}_{j}\operatorname{sen}\frac{2k\pi j}{2^{n-p}},\quad k=1,\ 2,\ \ldots,\ 2^{n-p-1}-1$$

donde $p=1,\ 2,\ \ldots,\ n-2$. Los coeficientes $a_j^{(p)}$ se encuentran por recurrencia según las fórmulas

$$a_j^{(p)}=a_j^{(p-1)}+a^{(p-1)}_{2^{n-p}+j},\quad j=0,\ 1,\ \ldots,\ 2^{n-p}-1, \qquad (64)$$
$$a^{(p)}_{2^{n-p}+j}=a_j^{(p-1)}-a^{(p-1)}_{2^{n-p}+j},\quad p=1,\ 2,\ \ldots,\ n$$

Poniendo $p = n - 1$ y $s = p = n - 1$ en (62), obtenemos

$$y_0 = a_0^{(n-1)} + a_1^{(n-1)} = a_0^{(n)},$$
$$y_{2^{n-1}} = a_0^{(n-1)} - a_1^{(n-1)} = a_1^{(n)}, \quad (65)$$
$$y_{2^{n-2}} = a_2^{(n-1)},$$

y de (63) para $p = n - 2$ hallaremos

$$\bar{y}_{2^{n-2}} = a_1^{(n-2)} - a_3^{(n-2)} = a_3^{(n-1)}. \quad (66)$$

Los restantes y_k y $\bar{y}_k$ se encuentran por las fórmulas

$$y_{2^{s-1}(2k-1)} = \sum_{j=0}^{2^{n-s}-1} a_{2^{n-s}+j}^{(s)} \cos \frac{(2k-1)\pi j}{2^{n-s}},$$

$$\bar{y}_{2^{s-1}(2k-1)} = \sum_{j=1}^{2^{n-s}-1} a_{2^{n-s}+j}^{(s)} \operatorname{sen} \frac{(2k-1)\pi j}{2^{n-s}},$$

$$k = 1, 2, \ldots, 2^{n-s-1}, \quad s = 1, 2, \ldots, n-2.$$

Realicemos aquí unos cambios para s fijo:

$$z_k^{(0)}(1) = y_{2^{s-1}(2k-1)}, \quad \bar{z}_k^{(0)}(1) = \bar{y}_{2^{s-1}(2k-1)},$$
$$k = 1, 2, \ldots, 2^{n-s-1}, \quad b_j^{(0)}(1) = a_{2^{n-s}+j}^{(s)},$$
$$j = 0, 1, \ldots, 2^{n-s} - 1,$$
$$l = n - s, \quad s = 1, 2, \ldots, n-2.$$

Esto nos conduce al cálculo de las sumas

$$z_k^{(0)}(1) = \sum_{j=0}^{2^l-1} b_j^{(0)}(1) \cos \frac{(2k-1)\pi j}{2^l};$$

$$\bar{z}_k^{(0)}(1) = \sum_{j=1}^{2^l-1} b_j^{(0)}(1) \operatorname{sen} \frac{(2k-1)\pi j}{2^l}, \quad (67)$$

$$k = 1, 2^{l-1}, \quad l = 2, 3, \ldots, n-1.$$

En la segunda etapa del algoritmo se calculan las sumas (67). Aquí, como en el algoritmo del punto 2, estas sumas se transforman mediante la separación de los sumandos con

índices j pares e impares y la utilización de las igualdades

$$\operatorname{sen} \frac{(2k-1)(2j-1)\pi}{2^{l-m+1}} + \operatorname{sen} \frac{(2k-1)2j\pi}{2^{l-m+1}} =$$
$$= 2 \cos \frac{(2k-1)\pi}{2^{l-m+1}} \operatorname{sen} \frac{(2k-1)(2j-1)\pi}{2^{l-m+1}},$$
$$\cos \frac{(2k-1)(2j-2)\pi}{2^{l-m+1}} + \cos \frac{(2k-1)2j\pi}{2^{l-m+1}} =$$
$$= 2 \cos \frac{(2k-1)\pi}{2^{l-m+1}} \cos \frac{(2k-1)(2j-1)\pi}{2^{l-m+1}}$$

para $m = 1, 2, \ldots$. Esto da las siguientes fórmulas recurrentes:

$$z_k^{(m-1)}(s) = z_k^{(m)}(2s) + \frac{1}{2 \cos \frac{(2k-1)\pi}{2^{l-m+1}}} z_k^{(m)}(2s-1),$$

$$z_{2^{l-m}-k+1}^{(m-1)}(s) = z_k^{(m)}(2s) - \frac{1}{2 \cos \frac{(2k-1)\pi}{2^{l-m+1}}} z_k^{(m)}(2s-1),$$

$$\bar{z}_k^{(m-1)}(s) = \bar{z}_k^{(m)}(2s) + \frac{1}{2 \cos \frac{(2k-1)\pi}{2^{l-m+1}}} \bar{z}_k^{(m)}(2s-1), \qquad (68)$$

$$\bar{z}_{2^{l-m}-k+1}^{(m-1)}(s) = -\bar{z}_k^{(m)}(2s) + \frac{1}{2 \cos \frac{(2k-1)\pi}{2^{l-m+1}}} \bar{z}_k^{(m)}(2s-1),$$

$$k = 1, 2, \ldots, 2^{l-m-1}, \quad s = 1, 2, \ldots, 2^{m-1},$$
$$m = 1, 2, \ldots, l-1$$

para el cálculo sucesivo de las sumas

$$z_k^{(m)}(s) = \sum_{j=0}^{2^{l-m}-1} b_j^{(m)}(s) \cos \frac{(2k-1)\pi j}{2^{l-m}},$$

$$\bar{z}_k^{(m)}(s) = \sum_{j=1}^{2^{l-m}-1} b_j^{(m)}(s) \operatorname{sen} \frac{(2k-1)\pi j}{2^{l-m}}, \qquad (69)$$

$$k = 1, 2, \ldots, 2^{l-m-1}, \quad s = 1, 2, \ldots, 2^m$$

con $m = 0, 1, \ldots, l-1$.

Los coeficientes $b_j^{(m)}(s)$ también se determinan por recurrencia para $s = 1, 2, \ldots, 2^{m-1}$, comenzando por los

$b_j^{(0)}(1)$ dados, según las fórmulas

$$b_j^{(m)}(2s-1)=b_{2j-1}^{(m-1)}(s)+b_{2j+1}^{(m-1)}(s),$$
$$j=1,\ 2,\ \ldots,\ 2^{l-m}-1,$$
$$b_0^{(m)}(2s-1)=b_1^{(m-1)}(s)-b_{2^{l-m+1}-1}^{(m-1)}(s), \tag{70}$$
$$b_j^{(m)}(2s)=b_{2j}^{(m-1)}(s),\quad j=0,\ 1,\ \ldots,\ 2^{l-m}-1,$$
$$s=1,\ 2,\ \ldots,\ 2^{m-1},\quad m=1,\ 2,\ \ldots,\ l-1.$$

Poniendo $m = l - 1$ en (69), obtendremos los valores iniciales para las relaciones (68).

$$z_1^{(l-1)}(s)=b_0^{(l-1)}(s),\quad \bar{z}_1^{(l-1)}(s)=b_1^{(l-1)}(s),$$
$$s=1,\ 2,\ \ldots,\ 2^{l-1}. \tag{71}$$

Así, el algoritmo del cálculo simultáneo de las sumas (57) y (58) se describe por las fórmulas (64)-(66), (68), (70) y (71). Notemos que, como en los algoritmos de los puntos 2 y 3, aquí en las relaciones (68) son posibles las sustituciones

$$z_k^{(m)}(s)=\operatorname{sen}\frac{(2k-1)\pi}{2^{l-m}}\,w_k^{(m)}(s),$$

$$\bar{z}_k^{(m)}(s)=\operatorname{sen}\frac{(2k-1)\pi}{2^{l-m}}\,\bar{w}_k^{(m)}(s),$$

los cuales permiten evitar la división por $2\cos\frac{(2k-1)\pi}{2^{l-m+1}}$.

Un cálculo elemental del número de operaciones aritméticas para el algoritmo construido nos da: $Q_+ = 3n/2\cdot 2^n - 1$ operaciones de suma y $Q_* = (n/2 - 3/2)\times 2^n + 2$ operaciones de multiplicación y en total $Q = (2\log_2 N - 3/2)\times \times N + 1$, $N = 2^n$.

De esta forma, el cálculo de los coeficientes de Fourier y la construcción de una función real periódica reticular por el algoritmo propuesto exigen $O(N \ln N)$ operaciones aritméticas.

5. Transformación de una función compleja periódica reticular. Examinemos ahora el problema 5 sobre el cálculo de los coeficientes de Fourier y la reconstrucción de una función compleja periódica reticular. En el punto 1 fue mostrado, que este problema se reduce al cálculo de las sumas (21),

las cuales en el caso $N = 2^n$ poseen la forma

$$y_k = \sum_{j=0}^{2^n-1} a_j^{(0)} e^{\frac{2k\pi j}{2^n} i}, \quad k = 0, 1, \ldots, 2^n - 1, \tag{72}$$

donde los $a_j^{(p)}$ son los números complejos.

El algoritmo para calcular las sumas (72) se construye igual que el algoritmo del cálculo de los coeficientes de Fourier de una función real periódica. En la primera etapa se agrupan primeramente los términos de las sumas (72) con índices j y $2^{n-1} + j$ para $j = 0, 1, \ldots, 2^{n-1} - 1$, luego con índices j y $2^{n-2} + j$ para $j = 0, 1, \ldots, 2^{n-2} - 1$, etc. Teniendo en cuenta la igualdad $e^{\pi k i} = (-1)^k$, obtenemos como resultado del p-ésimo paso las siguientes sumas:

$$y_{2^{s-1}(2k-1)} = \sum_{j=0}^{2^{n-s}-1} a_{2^{n-s}+j}^{(s)} e^{\frac{(2k-1)\pi j}{2^{n-s}} i},$$

$$k = 1, 2, \ldots, 2^{n-s}, \quad s = 1, 2, \ldots, p, \tag{73}$$

$$y_{2^p k} = \sum_{j=0}^{2^{n-p}-1} a_j^{(p)} e^{\frac{2k\pi j}{2^{n-s}} i}, \quad k = 0, 1, \ldots, 2^{n-p} - 1,$$

donde los coeficientes $a_j^{(p)}$ se encuentran por las fórmulas recurrentes (64).

Poniendo $s = p = n$ en (73), tendremos

$$y_0 = a_0^{(n)}, \quad y_{2^{n-1}} = a_1^{(n)} \tag{74}$$

y los restantes y_k se encuentran por las fórmulas

$$y_{2^{s-1}(2k-1)} = \sum_{j=0}^{2^{n-s}-1} a_{2^{n-s}+j}^{(s)} e^{\frac{(2k-1)\pi j}{2^{n-s}} i},$$

$$k = 1, 2, \ldots, 2^{n-s}, \quad s = 1, 2, \ldots, n-1.$$

Realicemos aquí unas sustituciones para j fijo, poniendo

$$z_k^{(0)}(1) = y_{2^{s-1}(2k-1)}, \quad k = 1, 2, \ldots, 2^{n-s},$$

$$b_j^{(0)}(1) = a_{2^{n-s}+j}^{(s)}, \quad j = 0, 1, \ldots, 2^{n-s} - 1,$$

$$l = n - s, \quad s = 1, 2, \ldots, n-1,$$

pasemos al cálculo de las sumas

$$z_k^{(0)}(1) = \sum_{j=0}^{2^l-1} b_j^{(0)}(1) e^{\frac{(2k-1)\pi j}{2^l} i}, \quad k = 1, 2, \ldots, 2^l \tag{75}$$

para $l = 1, 2, \ldots, n-1$.

La segunda etapa del algoritmo, la cual consiste en el cálculo de las sumas (75), se construye, como antes, por medio de la separación de los sumandos con índices j pares e impares al usar las igualdades

$$e^{\frac{(2k-1)(2j-2)\pi}{2^{l-m+1}}}+e^{\frac{(2k-1)2j\pi}{2^{l-m+1}}}=2\cos\frac{(2k-1)\pi}{2^{l-m+1}}\,e^{\frac{(2k-1)(2j-1)\pi}{2^{l-m+1}}}.$$

Tendremos las fórmulas recurrentes

$$\begin{aligned} z_k^{(m-1)}(s) &= z_k^{(m)}(2s)+\frac{1}{2\cos\frac{(2k-1)\pi}{2^{l-m+1}}}z_k^{(m)}(2s-1),\\ z_{2^{l-m}+k}^{(m-1)}(s) &= z_k^{(m)}(2s)-\frac{1}{2\cos\frac{(2k-1)\pi}{2^{l-m+1}}}z_k^{(m)}(2s-1), \end{aligned} \qquad (76)$$

$k=1,\ 2,\ \ldots,\ 2^{l-m},\quad s=1,\ 2,\ \ldots,\ 2^{m-1},$

$m=1,\ 2,\ \ldots,\ l-1$

para calcular las sumas

$$z_k^{(m)}(s)=\sum_{j=0}^{2^{l-m}-1} b_j^{(m)}(s)\,e^{\frac{(2k-1)\pi j}{2^{l-m}}i}, \qquad (77)$$

$k=1,\ 2,\ \ldots,\ 2^{l-m},\quad s=1,\ 2,\ \ldots,\ 2^m$

para $m=0,\ 1,\ \ldots,\ l-1$. Los coeficientes $b_j^{(m)}$ se calculan por las fórmulas recurrentes (70). Quedan por indicar los valores iniciales para (76). Poniendo $m=l-1$ en (77), obtendremos

$$\begin{aligned} z_1^{(l-1)}(s) &= b_0^{(l-1)}(s)+ib_1^{(l-1)}(s),\\ z_2^{(l-1)}(s) &= b_0^{(l-1)}(s)-ib_1^{(l-1)}(s),\quad s=1,2,\ldots,2^{l-1}. \end{aligned} \qquad (78)$$

De ese modo el algoritmo para calcular las sumas (72) se describe por las fórmulas (64), (70), (74), (76) y (78). Observemos, que el algoritmo construido no contiene (a excepción de la fórmula más simple (78)) operaciones de multiplicación de números complejos. Por eso en las fórmulas mostradas es fácil separar las partes real e imaginaria de las magnitudes calculadas. Esto es cómodo para la realización del algoritmo en una calculadora, la cual no posee la aritmética compleja. Más adelante, en las relaciones (76) puede

resultar útil la sustitución

$$z_k^{(m)}(s) = \operatorname{sen}\frac{(2k-1)\pi}{2^{l-m}} w_k^{(m)}(s).$$

Contemos ahora el número de operaciones aritméticas para el algoritmo construido. Obtenemos $Q_+ = (3n/2 - \frac{1}{2}) 2^n$ operaciones de suma de números complejos y $Q_* = (n/2 - 3/2) 2^n$ operaciones de multiplicación de un número complejo por un número real. Si expresamos estos valores en términos del número de operaciones sobre números reales, entonces obtendremos $Q_+ = (3n - 1) 2^n$ operaciones reales de suma y $Q_* = (n - 3) 2^n$ operaciones reales de producto, y en total $Q = (4 \log_2 N - 4) N$, $N = 2^n$ operaciones sobre números reales. Esta estimación supera en dos veces la obtenida en el punto 4 para el caso de una función real periódica reticular, lo cual es natural, ya que en el caso complejo examinado se trabaja con dos veces más números reales.

Con esto terminamos el exámen de los algoritmos de la transformación de Fourier discreta rápida y pasamos a su empleo para resolver de las ecuaciones elípticas reticulares.

§ 2. Resolución de problemas de diferencias por el método de Fourier

1. Problemas de diferencias en valores propios para el operador de Laplace en un rectángulo. En el § 5 del cap. I fueron examinados los problemas de contorno en valores propios para el operador de segunda derivada de diferencias, definido sobre una red uniforme en un intervalo. En el caso bidimensional los análogos de estos problemas son los problemas en valores propios para el operador de Laplace de diferencias, definido sobre una red rectangular uniforme en un rectángulo. Utilicemos el método de separación de variables para buscar los *valores propios* λ_k *y las funciones propias* $\mu_k(i, j)$ *del operador de Laplace de diferencias*

$$\Lambda = \Lambda_1 + \Lambda_2, \quad \Lambda_\alpha y = y_{\bar{x}_\alpha x_\alpha}, \quad \alpha = 1, 2.$$

Sea $\tilde{\omega} = \{x_{ij} = (ih_1, jh_2) \in \bar{G}, \ 0 \leqslant i \leqslant N_1, \ 0 \leqslant j \leqslant N_2, \ h_\alpha N_\alpha = l_\alpha, \ \alpha = 1, 2\}$ una red rectangular uniforme con pasos h_1 y h_2 definida en el rectángulo $\bar{G} =$

$= \{0 \leqslant x_\alpha \leqslant l_\alpha,\ \alpha = 1, 2\}$. Como es frecuente, designaremos mediante ω y γ los nodos interiores y de frontera respectivamente de la red $\bar{\omega}$.

El problema más simple en valores propios para el operador de Laplace en el caso de las condiciones de Dirichlet se plantea así: hallar aquellos valores del parámetro λ, para los cuales existen soluciones no triviales $y(x)$ del siguiente problema:

$$\begin{aligned} &\Lambda y(x) + \lambda y(x) = 0, \quad && x \in \omega, \\ &y(x) = 0, && x \in \gamma. \end{aligned} \tag{1}$$

Buscaremos la función propia $\mu_k(i, j)$ del problema (1) correspondiente al valor propio λ_k, en la forma

$$\mu_k(i, j) = \mu_{k_1}^{(1)}(i)\mu_{k_2}^{(2)}(j), \quad k = (k_1, k_2). \tag{2}$$

Sustituyamos en (1) la función $\mu_k(i, j)$ en lugar de $y(x_{ij}) = y(i, j)$. Ya que

$$\Lambda_1 y(i, j) = \frac{1}{h_1^2}[y(i+1, j) - 2y(i, j) + y(i-1, j)],$$

entonces el operador Λ_1 actúa solamente sobre una función reticular, dependiente del argumento i. Análogamente el operador Λ_2 actúa sobre una función que depende del argumento j. Por eso después de la sustitución de (2) en (1) tendremos

$$\mu_{k_2}^{(2)}(j)\Lambda_1\mu_{k_1}^{(1)}(i) + \mu_{k_1}^{(1)}(i)\Lambda_2\mu_{k_2}^{(2)}(j) + \lambda_k\mu_{k_1}^{(1)}(i)\mu_{k_2}^{(2)}(j) = 0, \tag{3}$$

para $1 \leqslant i \leqslant N_1 - 1$, $1 \leqslant j \leqslant N_2 - 1$, y además

$$\mu_{k_1}^{(1)}(0) = \mu_{k_1}^{(1)}(N_1) = 0, \quad \mu_{k_2}^{(2)}(0) = \mu_{k_2}^{(2)}(N_2) = 0. \tag{4}$$

De (3) encontramos, que

$$\frac{\Lambda_1\mu_{k_1}^{(1)}(i)}{\mu_{k_1}^{(1)}(i)} = -\frac{\Lambda_2\mu_{k_2}^{(2)}(j)}{\mu_{k_2}^{(2)}(j)} - \lambda_k. \tag{5}$$

Como el miembro izquierdo no depende de j, entonces tampoco depende de j el segundo miembro. Por otra parte, como el segundo miembro no depende de i, entonces tampoco depende de i el primer miembro. Por lo tanto, los miembros primero y segundo en (5) son constantes. Pongamos

$$\frac{\Lambda_1\mu_{k_1}^{(1)}(i)}{\mu_{k_1}^{(1)}(i)} = -\lambda_{k_1}^{(1)}, \quad \frac{\Lambda_2\mu_{k_2}^{(2)}(j)}{\mu_{k_2}^{(2)}(j)} = -\lambda_{k_2}^{(2)} \quad \lambda_k = \lambda_{k_1}^{(1)} + \lambda_{k_2}^{(2)} \tag{6}$$

y añadamos aquí las condiciones de contorno (4). Como resultado obtendremos los problemas unidimensionales reticulares en valores propios

$$\begin{aligned} &\Lambda_1\mu_{k_1}^{(1)}+\lambda_{k_1}^{(1)}\mu_{k_1}^{(1)}=0, \quad 1\leqslant i\leqslant N_1-1, \\ &\mu_{k_1}^{(1)}(0)=\mu_{k_1}^{(1)}(N_1)=0 \end{aligned} \tag{7}$$

y

$$\begin{aligned} &\Lambda_2\mu_{k_2}^{(2)}+\lambda_{k_2}^{(2)}\mu_{k_2}^{(2)}=0, \quad 1\leqslant j\leqslant N_2-1, \\ &\mu_{k_2}^{(2)}(0)=\mu_{k_2}^{(2)}(N_2)=0. \end{aligned} \tag{8}$$

Las soluciones de los problemas (7) y (8) fueron halladas por nosotros anteriormente en el § 5 del cap. I:

$$\lambda_{k_\alpha}^{(\alpha)}=\frac{4}{h_\alpha^2}\operatorname{sen}^2\frac{k_\alpha\pi}{2N_\alpha}=$$

$$=\frac{4}{h_\alpha^2}\operatorname{sen}^2\frac{k_\alpha\pi h_\alpha}{2l_\alpha}, \quad k_\alpha=1,\ 2,\ \ldots,\ N_\alpha-1,$$

$$\mu_{k_1}^{(1)}(i)=\sqrt{\frac{2}{l_1}}\operatorname{sen}\frac{k_1\pi i}{N_1}, \quad k_1=1,\ 2,\ \ldots,\ N_1-1,$$

$$\mu_{k_2}^{(2)}(j)=\sqrt{\frac{2}{l_2}}\operatorname{sen}\frac{k_2\pi j}{N_2}, \quad k_2=1,\ 2,\ \ldots,\ N_2-1.$$

Así, *están halladas las funciones propias y los valores propios del operador de Laplace de diferencias* Λ para el caso de las condiciones de contorno de Dirichlet

$$\begin{aligned} &\mu_k(i,\ j)=\mu_{k_1}^{(1)}(i)\,\mu_{k_2}^{(2)}(j)=\frac{2}{\sqrt{l_1l_2}}\operatorname{sen}\frac{k_1\pi i}{N_1}\operatorname{sen}\frac{k_2\pi j}{N_2}, \\ &0\leqslant i\leqslant N_1, \quad 0\leqslant j\leqslant N_2, \\ &\lambda_k=\lambda_{k_1}^{(1)}+\lambda_{k_2}^{(2)}=\sum_{\alpha=1}^{2}\frac{4}{h_\alpha^2}\operatorname{sen}^2\frac{k_\alpha\pi h_\alpha}{2l_\alpha}, \end{aligned} \tag{9}$$

donde $k_\alpha=1,\ 2,\ \ldots,\ N_\alpha-1$, $\alpha=1,\ 2$.

Observemos las propiedades fundamentales de las funciones propias y los valores propios hallados en (9). Introduzcamos el producto escalar de funciones reticulares, definidas sobre la red $\bar{\omega}$, de la siguiente forma:

$$(u,\ v)=\sum_{x\in\bar{\omega}}u(x)\,v(x)\,\hbar_1(x_1)\,\hbar_2(x_2),$$

$$\hbar_\alpha(x_\alpha)=\begin{cases}0{,}5h_\alpha, & x_\alpha=0,\ l_\alpha,\\ h_\alpha, & h_\alpha\leqslant x_\alpha\leqslant l_\alpha-h_\alpha.\end{cases}$$

Si designamos

$$(u, v)_{\bar{\omega}_\alpha} = \sum_{x_\alpha \in \bar{\omega}_\alpha} u(x)\, v(x)\, \hbar_\alpha(x_\alpha), \quad \alpha = 1, 2, \tag{10}$$

donde

$$\bar{\omega}_1 = \{x_1(i) = ih_1 \quad 0 \leqslant i \leqslant N_1\}, \quad \bar{\omega}_2 = \{x_2(j) = jh_2 \quad 0 \leqslant j \leqslant N_2\},$$

entonces es evidente, que $\bar{\omega} = \bar{\omega}_1 \times \bar{\omega}_2$ y $x_{ij} = (x_1(i), x_2(j))$, además

$$(u, v) = ((u, v)_{\bar{\omega}_1}, 1)_{\bar{\omega}_2} = ((u, v)_{\bar{\omega}_2}, 1)_{\bar{\omega}_1}. \tag{11}$$

Recordemos, que en el § 5 del cap. I, fue señalado, que las funciones reticulares $\mu_{k_1}^{(1)}(i)$ y $\mu_{k_2}^{(2)}(j)$ están ortonormalizadas en el sentido del producto escalar (10), es decir

$$(\mu_{k_\alpha}^{(\alpha)}, \mu_{m_\alpha}^{(\alpha)})_{\bar{\omega}_\alpha} = \delta_{k_\alpha, m_\alpha} = \begin{cases} 1, & k_\alpha = m_\alpha, \\ 0, & k_\alpha \neq m_\alpha. \end{cases}$$

Por eso de aquí y de (11) se deduce la ortonormalidad del sistema de funciones propias $\mu_k(i, j)$ definido por las fórmulas (9):

$$(\mu_k, \mu_m) = \delta_{k, m} = \begin{cases} 1, & k = m, \\ 0, & k \neq m, \quad k = (k_1, k_2), \quad m = (m_1, m_2). \end{cases}$$

Como el número de funciones propias $\mu_k(i, j) = \mu_{k_1, k_2}(i, j)$ es igual a $(N_1 - 1)(N_2 - 1)$ y coincide con el número de nodos interiores de la red $\bar{\omega}$, entonces cualquier función reticular $f(i, j)$, definida sobre ω (o sobre $\bar{\omega}$ y que se anule sobre γ), se puede representar en la siguiente forma:

$$f(i, j) = \sum_{k_1=1}^{N_1-1} \sum_{k_2=1}^{N_2-1} f_{k_1 k_2} \mu_{k_1}^{(1)}(i)\, \mu_{k_2}^{(2)}(j), \tag{12}$$

$$1 \leqslant i \leqslant N_1 - 1, \quad 1 \leqslant j \leqslant N_2 - 1,$$

donde los coeficientes de Fourier se definen de la siguiente forma:

$$f_k = f_{k_1 k_2} = (f, \mu_k) = \sum_{i=1}^{N_1-1} \sum_{j=1}^{N_2-1} f(i, j)\, \mu_{k_1}^{(1)}(i)\, \mu_{k_2}^{(2)}(j)\, h_1 h_2, \tag{13}$$

$$k = 1, 2, \ldots, N_1 - 1, \quad k_2 = 1, 2, \ldots, N_2 - 1.$$

Para los valores propios λ_k es válida la estimación

$$\lambda_{\min} = \lambda_1^{(1)} + \lambda_1^{(2)} \leqslant \lambda_k = \lambda_{k_1} + \lambda_{k_2} \leqslant \lambda_{N_1-1}^{(1)} + \lambda_{N_2-1}^{(2)} = \lambda_{\max},$$

donde

$$\lambda_{\min} = \sum_{\alpha=1}^{2} \frac{4}{h_\alpha^2} \operatorname{sen}^2 \frac{\pi h_\alpha}{2l_\alpha} \geqslant 8\left(\frac{1}{l_1^2} + \frac{1}{l_2^2}\right) > 0,$$

$$\lambda_{\max} = \sum_{\alpha=1}^{2} \frac{4}{h_\alpha^2} \cos^2 \frac{\pi h_\alpha}{2l_\alpha} < 4\left(\frac{1}{h_1^2} + \frac{1}{h_2^2}\right).$$

Examinemos ahora un ejemplo de un problema más complejo en valores propios para la ecuación de Laplace de diferencias. Supongamos que en los lados del rectángulo para $x_1 = 0$ y $x_1 = l_1$ de antemano están dadas las condiciones de Dirichlet, y para $x_2 = 0$ y $x_2 = l_2$ las condiciones de Neumann, es decir, hemos planteado el siguiente problema en valores propios:

$$\Lambda y(x) + \lambda y(x) = 0, \quad x \in \omega_1 \times \bar{\omega}_2,$$
$$y(x) = 0, \quad x_1 = 0, \quad x_2 = l_1. \tag{14}$$

Aquí $\Lambda = \Lambda_1 + \Lambda_2$, donde el operator Λ_1 fue definido anteriormente, y

$$\Lambda_2 y = \begin{cases} \dfrac{2}{h_2} y_{x_2}, & x_2 = 0, \\ y_{\bar{x}_2 x_2}, & h_2 \leqslant x_2 \leqslant l_2 - h_2, \\ -\dfrac{2}{h_2} y_{\bar{x}_2}, & x_2 = l_2. \end{cases} \tag{15}$$

Utilizando la definición de los operadores Λ_1 y Λ_2, el problema (14) se puede escribir en la siguiente forma:

$$y_{\bar{x}_1 x_1} + y_{\bar{x}_2 x_2} + \lambda y = 0, \quad x \in \omega,$$

$$\left.\begin{aligned} & y_{\bar{x}_1 x_1} + \frac{2}{h_2} y_{x_2} + \lambda y = 0, \quad x_2 = 0, \\ & y_{\bar{x}_1 x_1} - \frac{2}{h_2} y_{\bar{x}_2} + \lambda y = 0, \quad x_2 = l_2, \end{aligned}\right\} h_1 \leqslant x_1 \leqslant l_1 - h_1,$$

$$y(0, x_2) = y(l_1, x_2) = 0, \quad 0 \leqslant x_2 \leqslant l_2.$$

La solución del problema (14) se encuentra por el método de separación de variables. Sustituyendo en (14) en lugar de y la función reticular $\mu_k(i, j)$ de (2), obtendremos para $\mu_{k_1}^{(1)}(i)$ el problema (7), y para $\mu_{k_2}^{(2)}(j)$ tendremos el siguiente

problema de contorno:

$$\Lambda_2\mu_{k_2}^{(2)}+\lambda_{k_2}^{(2)}\mu_{k_2}^{(2)}=0, \quad 0\leqslant j\leqslant N_2$$

o en virtud de (15)

$$(\mu_{k_2}^{(2)})_{\bar{x}_2x_2}+\lambda_{k_2}^{(2)}\mu_{k_2}^{(2)}=0, \quad 1\leqslant j\leqslant N_2-1,$$
$$\frac{2}{h_2}(\mu_{k_2}^{(2)})_{x_2}+\lambda_{k_2}^{(2)}\mu_{k_2}^{(2)}=0, \quad j=0, \tag{16}$$
$$-\frac{2}{h_2}(\mu_{k_2}^{(2)})_{\bar{x}_2}+\lambda_{k_2}^{(2)}\mu_{k_2}^{(2)}=0, \quad j=N_2.$$

El problema (16) también fue resuelto por nosotros anteriormente en el § 5 del cap. I. La solución tiene la forma

$$\lambda_{k_2}^{(2)}=\frac{4}{h_2^2}\operatorname{sen}^2\frac{k_2\pi}{2N_2}=$$
$$=\frac{4}{h_2^2}\operatorname{sen}^2\frac{k_2\pi h_2}{2l_2}, \quad k_2=0,\ 1,\ \ldots,\ N_2,$$

$$\mu_{k_2}^{(2)}(j)=\begin{cases}\sqrt{\dfrac{2}{l_2}}\cos\dfrac{k_2\pi j}{N_2}, & 1\leqslant k_2\leqslant N_2-1,\\ \sqrt{\dfrac{1}{l_2}}\cos\dfrac{k_2\pi j}{N_2}, & k_2=0,\ N_2.\end{cases} \tag{17}$$

Así, hemos hallado la solución del problema (14), (15):

$$\mu_k(i,\ j)=\mu_{k_1}^{(1)}(i)\,\mu_{k_2}^{(2)}(j), \quad 0\leqslant i\leqslant N_1, \quad 0\leqslant j\leqslant N_2,$$
$$\lambda_k=\lambda_{k_1}^{(1)}+\lambda_{k_2}^{(2)}, \quad 1\leqslant k_1\leqslant N_1-1, \quad 0\leqslant k_2\leqslant N_2,$$

donde $\lambda_{k_1}^{(1)}$ y $\mu_{k_1}^{(1)}(i)$ están definidos más arriba, y $\lambda_{k_2}^{(2)}$ junto con $\mu_{k_2}^{(2)}(j)$ están definidos en (17).

Análogamente se resuelven los problemas en valores propios para el operador de Laplace de diferencias en un rectángulo y en el caso de otras combinaciones de condiciones de contorno sobre los lados del rectángulo G. El método de separación de variables permite reducir estos problemas a problemas unidimensionales, cuyas soluciones fueron obtenidas en el § 5 del cap. I. La generalización al caso multidimensional es evidente. Recordemos, que la solución analítica en forma de senos y cosenos de los correspondientes problemas unidimensionales fue obtenida en el § 5 del cap. I solamente para condiciones de contorno de primero y segundo género, sus combinaciones, y también para el caso del problema de contorno periódico. Por eso, si en los lados

del rectángulo (o en las caras del paralelepípedo rectangular en el caso tridimensional) se dan las condiciones de contorno enumeradas, entonces las funciones propias del operador de Laplace de diferencias se representan en la forma de un producto de senos y cosenos.

2. Ecuación de Poisson en un rectángulo. Desarrollo en serie doble. Examinemos ahora el método de separación de variables aplicado a la resolución *del problema de Dirichlet de diferencias para la ecuación de Poisson* sobre una red uniforme en un rectángulo:

$$\begin{gathered} \Lambda y = -\varphi(x), \quad x \in \omega, \quad y(x) = g(x), \quad x \in \gamma, \\ \Lambda = \Lambda_1 + \Lambda_2, \quad \Lambda_\alpha y = y_{\bar{x}_\alpha x_\alpha}, \quad \alpha = 1, 2. \end{gathered} \qquad (18)$$

Primeramente reduciremos el problema (18) a un problema con condiciones de contorno homogéneas mediante el cambio del segundo miembro de la ecuación en los nodos fronterizos. El método estandarizado de dicha transformación consiste en el traslado de las magnitudes conocidas al miembro derecho de la ecuación escrita en un nodo fronterizo. Por ejemplo, si $x = (h_1, h_2) \in \omega$, entonces la ecuación de Poisson en este punto se escribe en la siguiente forma:

$$\frac{1}{h_1^2}[y(0, h_2) - 2y(h_1, h_2) + y(2h_1, h_2)] + \\ + \frac{1}{h_2^2}[y(h_1, 0) - 2y(h_1, h_2) + y(h_1, 2h_2)] = -\varphi(h_1, h_2)].$$

Como $y(0, h_2) = g(0, h_2)$, $y(h_1, 0) = g(h_1, 0)$, entonces trasladando estas magnitudes del primer miembro al segundo miembro de la ecuación, tendremos

$$\frac{1}{h_1^2}[-2y(h_1, h_2) + y(2h_1, h_2)] + \\ + \frac{1}{h_2^2}[-2y(h_1, h_2) + y(h_1 2h_2)] = \\ -\left[\varphi(h_1, h_2) + \frac{1}{h_1^2} g(0, h_2) + \frac{1}{h_2^2} g(h_1, 0)\right].$$

Realizando una transformación semejante para cada punto fronterizo, obtendremos ecuaciones en diferencias, que no contienen los valores $y(x)$ sobre γ en el miembro izquierdo. Los segundos miembros de las ecuaciones para los nodos fronterizos se diferencian del miembro derecho $\varphi(x)$. Si designamos mediante $f(x)$ el segundo miembro constru-

ido, entonces se determina por las fórmulas

$$f(x)=\varphi(x)+\frac{1}{h_1^2}\varphi_1(x)+\frac{1}{h_2^2}\varphi_2(x), \quad x\in\omega, \qquad (19)$$

donde

$$\varphi_1(x)=\begin{cases} g(0, x_2), & x_1=h_1, \\ 0, & 2h_1\leqslant x_1\leqslant l_1-2h_1, \\ g(l_1, x_2), & x_1=l_2, \end{cases}$$

$$\varphi_2(x)=\begin{cases} g(x_1, 0), & x_2=h_2, \\ 0, & 2h_2\leqslant x_2\leqslant l_2-2h_2, \\ g(x_1, l_2), & x_2=l_2. \end{cases}$$

El primer miembro de las ecuaciones transformadas se distingue de la escritura del operador de Laplace de diferencias para los nodos fronterizos. Sin embargo si ponemos $y(x) = u(x)$, $x\in\omega$, $u(x)=0$, $x\in\gamma$, entonces en todos los nodos de la red ω las ecuaciones se escribirán igual:

$$\Lambda u=-f(x), \quad x\in\omega, \quad u(x)=0, \quad x\in\gamma. \qquad (20)$$

Ya que $u(x)$ coincide con $y(x)$ para $x\in\omega$, entonces es suficiente hallar la solución del problema (20).

Hallemos la solución del problema (20). Como la función $u(x)$ se anula sobre γ, entonces en virtud de lo dicho más arriba ella puede ser representada en la forma de un desarrollo según las funciones propias $\mu_k(i, j)$ del operador de Laplace

$$u(i, j)=\sum_{k_1=1}^{N_1-1}\sum_{k_2=1}^{N_2-1} u_{k_1k_2}\mu_{k_1}^{(1)}(i)\,\mu_{k_2}^{(2)}(j), \qquad (21)$$

que es válido para $0\leqslant i\leqslant N_1$, $0\leqslant j\leqslant N_2$. Además, la función reticular $f(x)$ definida sobre ω, también admite la representación

$$f(i, j)=\sum_{k_1=1}^{N_1-1}\sum_{k_2=1}^{N_2-1} f_{k_1k_2}\mu_{k_1}^{(1)}(i)\,\mu_{k_2}^{(2)}(j) \qquad (22)$$

para $1\leqslant i\leqslant N_1-1$, $1\leqslant j\leqslant N_2-1$, donde los coeficientes de Fourier $f_{k_1k_2}$ están definidas en (13). Ya que $\mu_k(i, j)=\mu_{k_1}^{(1)}(i)\mu_{k_2}^{(2)}(j)$ es una función propia del operador de Laplace, correspodiente al valor propio λ_k, es decir

$$\Lambda\mu_k+\lambda_k\mu_k=0, \quad x\in\omega, \quad \lambda_{k_1}^{(1)}+\lambda_{k_2}^{(2)}=\lambda_k,$$

entonces después de la sustitución de (21) y (22) en la ecuación (20) tendremos

$$\Lambda u = \sum_{k_1=1}^{N_1-1} \sum_{k_2=1}^{N_2-1} (\lambda_{k_1}^{(1)} + \lambda_{k_2}^{(2)}) u_{k_1k_2} \mu_{k_1}^{(1)}(i) \mu_{k_2}^{(2)}(j) = -f(i, j) =$$

$$= -\sum_{k_1=1}^{N_1-1} \sum_{k_2=1}^{N_2-1} f_{k_1k_2} \mu_{k_1}^{(1)}(i) \mu_{k_2}^{(2)}(j),$$

$1 \leqslant i \leqslant N_1 - 1, \quad 1 \leqslant j \leqslant N_2 - 1.$

Utilizando la ortonormalidad de las funciones propias $\mu_k(i, j)$, de aquí obtenemos las siguientes igualdades:

$$u_{k_1k_2} = \frac{f_{k_1k_2}}{\lambda_{k_1}^{(1)} + \lambda_{k_2}^{(2)}}, \quad 1 \leqslant k_1 \leqslant N_1 - 1, \quad 1 \leqslant k_2 \leqslant N_2 - 1.$$

Sustituyendo esta expresión en (21), obtendremos la siguiente representación para la solución del problema (20):

$$u(i, j) = \sum_{k_1=1}^{N_1-1} \sum_{k_2=1}^{N_2-1} \frac{f_{k_1k_2}}{\lambda_{k_1}^{(1)} + \lambda_{k_2}^{(2)}} \mu_{k_1}^{(1)}(i) \mu_{k_2}^{(2)}(j),$$

$$0 \leqslant i \leqslant N_1, \quad 0 \leqslant j \leqslant N_2. \tag{23}$$

De esta forma, las fórmulas (13) y (23) dan la solución del problema (20). Analicémoslas desde el punto de vista computacional. Durante el cálculo de la solución $u(i, j)$ según las fórmulas (13) y (20), donde $\mu_k(i, j) = \mu_{k_1}^{(1)}(i) \mu_{k_2}^{(2)}(j)$ y $\lambda_k = \lambda_{k_1}^{(1)} + \lambda_{k_2}^{(2)}$ están definidos en (9), es útil introducir tres magnitudes auxiliares: $\varphi_{k_2}(i)$, $\varphi_{k_1k_2}$ y $u_{k_2}(i)$. Entonces el proceso numérico se puede organizar de la siguiente forma:

$$\varphi_{k_2}(i) = \sum_{j=1}^{N_2-1} f(i, j) \operatorname{sen} \frac{k_2 \pi j}{N_2},$$

$$1 \leqslant k_2 \leqslant N_2 - 1, \quad 1 \leqslant i \leqslant N_1 - 1, \tag{24}$$

$$\varphi_{k_1k_2} = \sum_{i=1}^{N_1-1} \varphi_{k_2}(i) \operatorname{sen} \frac{k_1 \pi i}{N_1},$$

$$1 \leqslant k_1 \leqslant N_1 - 1, \quad 1 \leqslant k_2 \leqslant N_2 - 1, \tag{25}$$

$$u_{k_2}(i) = \sum_{k_1=1}^{N_1-1} \frac{\varphi_{k_1k_2}}{\lambda_{k_1}^{(1)} + \lambda_{k_2}^{(2)}} \operatorname{sen} \frac{k_1 \pi i}{N_1},$$

$$1 \leqslant i \leqslant N_1 - 1, \quad 1 \leqslant k_2 \leqslant N_2 - 1, \tag{26}$$

$$u(i,\ j)=\frac{4}{N_1N_2}\sum_{k_2=1}^{N_2-1} u_{k_2}(i)\ \mathrm{sen}\,\frac{k_2\pi j}{N_2},$$
$$1\leqslant j\leqslant N_2-1,\quad 1\leqslant i\leqslant N_1-1. \qquad (27)$$

Contemos el número de operaciones aritméticas para el algoritmo (24) — (27), suponiendo que las magnitudes $(\lambda_{k_1}^{(1)}+ +\lambda_{k_2}^{(2)})^{-1}$ están dadas, y las sumas (24) — (27) se calculan utilizando el algoritmo de la transformación de Fourier rápida, expuesto en el punto 2 del § 1. Para aplicar el algoritmo indicado, es necesario suponer que N_1 y N_2 son múltiplos de 2: $N_1 = 2^n$, $N_2 = 2^m$.

Recordemos, que las sumas del tipo

$$y_k=\sum_{j=1}^{2^n-1} a_j\ \mathrm{sen}\,\frac{k\pi j}{2^n},\qquad k=1,\ 2,\ \ldots,\ 2^n-1,$$

se calculan con un gasto de $Q_+ = (3/2n - 2)\ 2^n - n + 2$ operaciones de suma y resta y $Q_* = (n/2-1)\ 2^n + 1$ operaciones de multiplicación, si se utiliza el algoritmo del punto 2 del § 1.

Un cálculo elemental nos da los siguientes gastos de operaciones aritméticas para el cómputo de la solución $u\ (i,\ j)$ según las fórmulas (24) — (27):

$$Q_+ = (N_1N_2 - N_1 - N_2)[3\log_2(N_1N_2) - 8] + \\ + (N_1+2)\log_2 N_2 + (N_2+2)\log_2 N_1 - 8$$

operaciones de suma y resta y

$$Q_* = (N_1N_2 - N_1 - N_2)[\log_2(N_1N_2) - 2] + \\ + N_1\log_2 N_2 + N_2\log_2 N_1 - 2$$

operaciones de multiplicación. Si no hacemos distinción entre las operaciones aritméticas, entonces para $N_1 = N_2 = N = 2^n$ el número total de operaciones para el algoritmo (24) — (27) es igual a

$$Q = (N^2 - 1{,}5\ N)\ (8\log_2 N - 10) + 5N + 4\log_2 N - 10.$$

De esta forma, el método descrito para resolver el problema (20) puede ser realizado con un gasto $O\ (N^2\log_2 N)$ de operaciones aritméticas. Tal tipo de estimación para el número de operaciones aritméticas lo posee el método de reducción completa y fue examinado en el capítulo III. La comparación de estas estimaciones muestra, que el algoritmo da-

do del método de separación de variables exige aproximadamente 1,5 veces más operaciones, que el método de reducción completa.

Observemos, que se puede construir un algoritmo, análogo al propuesto más arriba, en el caso cuando sobre los lados del rectángulo se da cualquier combinación de condiciones de contorno de primero o segundo género y condiciones de periodicidad, para las cuales el problema de diferencias sea no degenerado. Es necesario solamente sustituir las respectivas funciones propias y valores propios en (13) y (23), coordinar los límites de sumación con el tipo de las condiciones de contorno y utilizar el correspondiente algoritmo de la transformación de Fourier rápida del § 1 para calcular las sumas que aparezcan en este caso. La estimación del número de operaciones será del mismo tipo que para el caso del problema de Dirichlet examinado más arriba.

Nosotros hemos descrito la variante más simple del método de separación de variables. Si se exige resolver un problema de contorno de diferencias más general, por ejemplo la ecuación de Poisson en sistemas de coordenadas polares o cilíndricos, con condiciones de contorno que admiten separación de variables, entonces de nuevo se pueden utilizar los desarrollos (21) y (22). Pero en este caso al menos una de las funciones propias $\mu_{k_1}^{(1)}(i)$ y $\mu_{k_2}^{(2)}(j)$ es distinta de seno o coseno. Esto impide aprovechar el algoritmo de la transformación de Fourier rápida para el cálculo de todas las sumas necesarias. Por eso para tales problemas el número de operaciones aritméticas será del mismo orden, que en el caso del cálculo directo de las sumas sin tener en cuenta el tipo de las funciones propias $\mu_{k_1}^{(1)}(i)$ y $\mu_{k_2}^{(2)}(j)$, es decir $O(N^3)$.

Por consiguiente, es necesario modificar el método construido, para que en el caso cuando al menos una de las funciones $\mu_{k_1}^{(1)}(i)$ o $\mu_{k_2}^{(2)}(j)$ sea seno o coseno, el número de operaciones aritméticas se quede siendo una magnitud de orden $O(N^2 \log_2 N)$. Naturalmente, los problemas examinados en este punto pueden ser resueltos también por el método modificado y como resultará más abajo con un menor número de operaciones aritméticas. Dicho método —desarrollo en serie de aspecto singular— será construido en el punto 3. Desde el punto de vista numérico él se diferencia del método construido aquí en que no se calculan las dos sumas de (24)—(27), y en su lugar se resuelve una serie de problemas de contorno para ecuaciones en diferencias tripuntuales.

3. Desarrollo en serie de aspecto singular. Regresemos al problema (20):

$$\Lambda u = -f(x), \quad x \in \omega, \quad u(x) = 0, \quad x \in \gamma,$$
$$\Lambda = \Lambda_1 + \Lambda_2, \quad \Lambda_\alpha u = u_{\bar{x}_\alpha x_\alpha}, \quad \alpha = 1, 2. \tag{28}$$

Examinaremos la función buscada $u(x_{ij}) = u(i, j)$ y la función dada $f(i, j)$ para i fijo, $0 \leqslant i \leqslant N_1$, como funciones reticulares del argumento j. Ya que $u(i, j)$ se anula para $j = 0$ y $j = N_2$, y $f(i, j)$ está definida para $1 \leqslant j \leqslant N_2 - 1$, entonces ellas pueden ser representadas en forma de suma según las funciones propias $\mu_{k_2}^{(2)}(j)$ del operador de diferencias Λ_2:

$$u(i, j) = \sum_{k_2=1}^{N_2-1} u_{k_2}(i)\, \mu_{k_2}^{(2)}(j), \quad 0 \leqslant j \leqslant N_2, \quad 0 \leqslant i \leqslant N_1, \tag{29}$$

$$f(i, j) = \sum_{k_2=1}^{N_2-1} f_{k_2}(i)\, \mu_{k_2}^{(2)}(j), \quad 1 \leqslant j \leqslant N_2 - 1,$$
$$1 \leqslant i \leqslant N_1 - 1, \tag{30}$$

donde

$$\mu_{k_2}^{(2)}(j) = \frac{2}{l_2} \operatorname{sen} \frac{k_2 \pi j}{N_2}, \quad k_2 = 1, 2, \ldots, N_2 - 1. \tag{31}$$

Sustituyamos las expresiones (29) y (30) en (28) y tengamos en cuenta las igualdades

$$\begin{gathered} \Lambda_2 \mu_{k_2}^{(2)} + \lambda_{k_2}^{(2)} \mu_{k_2}^{(2)} = 0, \quad 1 \leqslant j \leqslant N_2 - 1, \\ \mu_{k_2}^{(2)}(0) = \mu_{k_2}^{(2)}(N_2) = 0. \end{gathered} \tag{32}$$

Como resultado obtendremos

$$\sum_{k_2=1}^{N_2-1} [\Lambda_1 u_{k_2}(i) - \lambda_{k_2}^{(2)} u_{k_2}(i) + f_{k_2}(i)]\, \mu_{k_2}^{(2)}(j) = 0$$

para $1 \leqslant i \leqslant N_1 - 1$, $1 \leqslant j \leqslant N_2 - 1$, y además $u_{k_2}(0) = u_{k_2}(N_1) = 0$, $k_2 = 1, 2, \ldots, N_2 - 1$.

De aquí, en virtud de la ortogonalidad del sistema de funciones propias $\mu_{k_2}^{(2)}(j)$, obtenemos una serie de problemas de contorno para determinar las funciones $u_{k_2}(i)$, $k_2 = 1, 2, \ldots, N_2 - 1$:

$$\begin{gathered} \Lambda_1 u_{k_2}(i) - \lambda_{k_2}^{(2)} u_{k_2}(i) = -f_{k_2}(i), \\ 1 \leqslant i \leqslant N_1 - 1, \\ u_{k_2}(0) = u_{k_2}(N_1) = 0, \end{gathered} \tag{33}$$

Los valores propios $\lambda_{k_2}^{(2)}$ del problema (32) son conocidos

$$\lambda_{k_2}^{(2)} = \frac{4}{h_2^2} \operatorname{sen}^2 \frac{k_2\pi}{2N_2}, \quad k_2 = 1, 2, \ldots, N_2 - 1, \tag{34}$$

y los coeficientes de Fourier $f_{k_2}(i)$ para cada $1 \leqslant i \leqslant N_1 - 1$ se calculan por las fórmulas

$$f_{k_2}(i) = (f, \mu_{k_2}^{(2)})_{\bar{\omega}_2} = \sum_{j=1}^{N_2-1} h_2 f(i, j) \mu_{k_2}^{(2)}(j),$$
$$1 \leqslant k_2 \leqslant N_2 - 1. \tag{35}$$

De esta forma, hemos hallado las fórmulas (29), (31), y (33)—(35) que describen completamente el método de resolución del problema (20). Por las fórmulas (35) se encuentran para $1 \leqslant i \leqslant N_1 - 1$ las funciones $f_{k_2}(i)$, a continuación se resuelven los problemas (33) con $1 \leqslant k_2 \leqslant N_2 - 1$ para determinar las funciones $u_{k_2}(i)$, y por las fórmulas (29) se calcula la solución buscada $u(i, j)$.

Examinemos ahora el algoritmo que realiza el método indicado. En lugar de $u_{k_2}(i)$ y $f_{k_2}(i)$ es cómodo introducir nuevas funciones auxiliares $v_{k_2}(i)$ y $\varphi_{k_2}(i)$ según las fórmulas

$$u_{k_2}(i) = \frac{\sqrt{2l_2}}{N_2} v_{k_2}(i), \quad f_{k_2}(i) = \frac{\sqrt{2l_2}}{N_2} \varphi_{k_2}(i). \tag{36}$$

Sustituyamos (31) y (36) en (29), (33) y (35), teniendo en cuenta que $h_2 N_2 = l_2$, y anotemos el operador de diferencias Λ_1 por puntos. Como resultado obtenemos

$$\varphi_{k_2}(i) = \sum_{j=1}^{N_2-1} f(i, j) \operatorname{sen} \frac{k_2 \pi j}{N_2}, \quad \left. \begin{array}{l} 1 \leqslant k_2 \leqslant N_2 - 1, \\ 1 \leqslant i \leqslant N_1 - 1, \end{array} \right\} \tag{37}$$

$$\left. \begin{array}{l} -v_{k_2}(i-1) + (2 + h_1^2 \lambda_{k_2}^{(2)}) v_{k_2}(i) - v_{k_2}(i+1) = h_1^2 \varphi_{k_2}(i), \\ 1 \leqslant i \leqslant N_1 - 1, \ v_{k_2}(0) = v_{k_2}(N_1) = 0, \ 1 \leqslant k_2 \leqslant N_2 - 1, \end{array} \right\} \tag{38}$$

$$u(i, j) = \frac{2}{N_2} \sum_{k_2=1}^{N_2-1} v_{k_2}(i) \operatorname{sen} \frac{k_2 \pi j}{N_2}, \quad \left. \begin{array}{l} 1 \leqslant j \leqslant N_2 - 1, \\ 1 \leqslant i \leqslant N_1 - 1, \end{array} \right\} \tag{39}$$

donde $\lambda_{k_2}^{(2)}$ está definido en (34).

Evidentemente, las sumas (37) y (39) se deben calcular, utilizando el algoritmo de la transformación de Fourier discreta rápida expuesto en el punto 2 del § 1. Para la resolución de los problemas de contorno tripuntuales (38) es oportuno utilizar el algoritmo de factorización, construido en

el § 1 del cap. II. Para el problema (38) el algoritmo de factorización se describe mediante las fórmulas

$$\alpha_{i+1}=\frac{1}{c_{k_2}-\alpha_i}, \quad 1\leqslant i\leqslant N_1-1, \quad \alpha_1=0,$$

$$\beta_{i+1}=[h_1^2\varphi_{k_2}(i)+\beta_i]\,\alpha_{i+1}, \quad 1\leqslant i\leqslant N_1-1, \quad \beta_1=0, \qquad (40)$$

$$v_{k_2}(i)=\alpha_{i+1}v_{k_2}(i+1)+\beta_{i+1}, \quad 1\leqslant i\leqslant N_1-1, \quad v_{k_2}(N_1)=0,$$

donde

$$c_{k_2}=2+h_1^2\lambda_{k_2}^{(2)} \text{ y } k_2=1,\ 2,\ \ldots,\ N_2-1.$$

Comparemos las fórmulas (37), (39) y (40) con las obtenidas anteriormente (24)—(27) para el método del desarrollo en serie doble. Aquí en lugar del cálculo de las dos sumas (25) y (26) nosotros resolvemos la serie de problemas de contorno (38) por el método de factorización (40). Por eso en el cómputo de las sumas (37) y (39) se gastarán aproximadamente dos veces menos operaciones aritméticas, que para el algoritmo (24)—(27). Los gastos complementarios en la resolución de los problemas (38) constituyen, evidentemente, $O\ (N_1N_2)$ operaciones, lo cual no influye en el término principal en la estimación del número de operaciones aritméticas del algoritmo (37), (39), (40). Citemos las estimaciones exactas del número de operaciones aritméticas para este algoritmo. Tenemos (para $N_2 = 2^m$) $Q_\pm = [(3\ \log_2 N_2 - 1)N_2 - 2\ \log_2 N_2 + 1](N_1 - 1)$ operaciones de suma y resta, $Q_* = [(\log_2 N_2 + 2)N_2 - 2]\ (N_1 - 1)$ operaciones de multiplicación y $Q_1 = (N_1 - 1)\ (N_2 - 1)$ operaciones de división, y en total para $N_1 = N_2 = N = 2^n$ el número de operaciones es igual a

$$Q = (N^2 - 1{,}5\ N)\ (4\ \log_2 N + 2) - N + 2\ \log_2 N + 2.$$

Nosotros hemos examinado el método del desarrollo en serie de aspecto singular en el ejemplo del problema de Dirichlet de diferencias para la ecuación de Poisson. El momento esencial está en que las funciones propias del operador de diferencias Λ_2 admiten la utilización del algoritmo de la transformación de Fourier rápida para el cálculo de las sumas correspondientes. Esta posibilidad tendrá lugar además para el caso, cuando sobre los lados $x_2 = 0$ y $x_2 = l_2$ del rectángulo $\overline{G}$ están prefijadas condiciones de contorno del rectángulo G están prefijadas condiciones de contorno de segundo género o combinaciones de condiciones de primero y segundo género en lugar de las condiciones de contorno

de primer género, y también para el caso de las condiciones periódicas.

Examinemos como un ejemplo el siguiente problema de contorno para la ecuación de Poisson:

$$
\begin{aligned}
&u_{\bar{x}_1x_1}+u_{\bar{x}_2x_2}=-\varphi(x), \quad x\in\omega,\\
&u(x)=0, \quad x_1=0, \quad l_1, \quad 0\leqslant x_2\leqslant l_2,\\
&u_{\bar{x}_1x_1}+\frac{2}{h_2}u_{x_2}=-\varphi(x)-\frac{2}{h_2}g_{-2}(x), \quad x_2=0, \qquad (41)\\
&u_{\bar{x}_1x_1}-\frac{2}{h_2}u_{\bar{x}_2}=-\varphi(x)-\frac{2}{h_2}g_{+2}(x), \quad x_2=l_2,\\
&h_1\leqslant x_1\leqslant l_2-h_1.
\end{aligned}
$$

El esquema (41) es una aproximación de diferencias del problema

$$
\begin{aligned}
&\frac{\partial^2u}{\partial x_1^2}+\frac{\partial^2u}{\partial x_2^2}=-\varphi(x), \quad x\in G,\\
&u(x)=0, \quad x_1=0, \quad l_1, \quad 0\leqslant x_2\leqslant l_2,\\
&\frac{\partial u}{\partial x_2}=-g_{-2}(x), \quad x_2=0,\\
&-\frac{\partial u}{\partial x_2}=-g_{+2}(x), \quad x_2=l_2, \quad 0\leqslant x_1\leqslant l_1.
\end{aligned}
$$

Escribamos el problema (41) en otra forma, introduciendo las notaciones:

$$
\Lambda_2u=\begin{cases}\dfrac{2}{h_2}u_{x_2}, & x_2=0,\\ u_{\bar{x}_2x_2}, & h_2\leqslant x_2\leqslant l_2-h_2,\\ -\dfrac{2}{h_2}u_{\bar{x}_2}, & x_2=l_2,\end{cases}
$$

$$
\varphi_2(x)=\begin{cases}\dfrac{2}{h_2}g_{-2}(x), & x_2=0,\\ 0, & h_2\leqslant x_2\leqslant l_2-h_2,\\ \dfrac{2}{h_2}g_{+2}(x), & x_2=l_2,\end{cases}
$$

$$
f(x)=\varphi(x)+\varphi_2(x), \qquad \Lambda_1u=u_{\bar{x}_1x_1}
$$

para $h_1\leqslant x_1\leqslant l_1-h_1$, $0\leqslant x_2\leqslant l_2$.

En las nuevas notaciones el problema (41) se escribe en la forma

$$\Lambda u = (\Lambda_1 + \Lambda_2)u = -f(x), \quad h_1 \leqslant x_1 \leqslant l_1 - h_1, \quad 0 \leqslant x_2 \leqslant l_2, \tag{42}$$
$$u(x) = 0, \quad x_1 = 0, \ l_1, \quad 0 \leqslant x_2 \leqslant l_2.$$

Desarrollando $u(i, j)$ y $f(i, j)$ en sumas según las funciones propias del operador Λ_2, tendremos

$$u(i, j) = \sum_{k_2=0}^{N_2} u_{k_2}(i)\,\mu_{k_2}^{(2)}(j), \quad 0 \leqslant j \leqslant N_2, \quad 0 \leqslant i \leqslant N_1,$$
$$f(i, j) = \sum_{k_2=0}^{N_2} f_{k_2}(i)\,\mu_{k_2}^{(2)}(j), \quad 0 \leqslant j \leqslant N_2, \quad 1 \leqslant i \leqslant N_1 - 1, \tag{43}$$

donde

$$\mu_{k_2}^{(2)}(j) = \begin{cases} \sqrt{\dfrac{1}{l_2}} \cos \dfrac{k_2 \pi j}{N_2}, & k_2 = 0, \ N_2, \\ \sqrt{\dfrac{2}{l_2}} \cos \dfrac{k_2 \pi j}{N_2}, & 1 \leqslant k_2 \leqslant N_2 - 1 \end{cases}$$

es la función propia del operador Λ_2, correspondiente al valor propio

$$\lambda_{k_2}^{(2)} = \frac{4}{h_2^2} \operatorname{sen}^2 \frac{k_2 \pi}{2N_2}, \quad k_2 = 0, \ 1, \ \ldots, \ N_2. \tag{44}$$

El coeficiente de Fourier $f_{k_2}(i)$ se calcula para cada $1 \leqslant i \leqslant N_1 - 1$ por las fórmulas

$$f_{k_2}(i) = \sum_{j=1}^{N_2-1} h_2 f(i, j)\,\mu_{k_2}^{(2)}(j) + 0{,}5 h_2 [f(i, 0)\,\mu_{k_2}^{(2)}(0) + f(i, N_2)\,\mu_{k_2}^{(2)}(N_2)].$$

Sustituyendo (43) en (42), obtendremos el siguiente análogo de las fórmulas (37)—(39) para el problema examinado:

$$\varphi_{k_2}(i) = \sum_{j=0}^{N_2} \rho_j f(i, j) \cos \frac{k_2 \pi j}{N_2},$$
$$0 \leqslant k_2 \leqslant N_2 \quad 1 \leqslant i \leqslant N_1 - 1,$$
$$-v_{k_2}(i-1) + (2 + h_1^2 \lambda_{k_2}^{(2)})\, v_{k_2}(i) - v_{k_2}(i+1) = h_1^2 \varphi_{k_2}(i),$$
$$1 \leqslant i \leqslant N_1 - 1, \quad v_{k_2}(0) = v_{k_2}(N_1) = 0, \quad 0 \leqslant k_2 \leqslant N_2,$$
$$u(i, j) = \frac{2}{N_2} \sum_{k_2=0}^{N_2} \rho_{k_2} v_{k_2}(i) \cos \frac{k_2 \pi j}{N_2},$$
$$0 \leqslant j \leqslant N_2, \quad 1 \leqslant i \leqslant N_1 - 1,$$

donde $\lambda_{k_2}^{(2)}$ está definido en (44), y

$$\rho_j = \begin{cases} 0{,}5, & j=0,\ N_2, \\ 1, & 1 \leqslant j \leqslant N_2-1. \end{cases}$$

Citamos un estimado del número de operaciones para el algoritmo construido siendo $N_1 = N_2 = N = 2^n$: $Q_{\pm} = [(3 \log_2 N_2 - 1) N_2 + 2 \log_2 N_2 + 7] (N_1 - 1)$ operaciones de suma y resta, $Q_* = [(\log_2 N_2 + 2) N_2 + 10] (N_1 - 1)$ operaciones de multiplicación y $Q_I = (N_2 + 1)(N_1 - 1)$ operaciones de división, y en total

$$Q = \left(N^2 - \frac{N}{2}\right) (4 \log_2 N + 2) + 17N - 2 \log_2 N - 18.$$

A continuación, como en el método de desarrollo en serie de aspecto singular no se utilizan las funciones propias del operador de diferencias Λ_1 y la única exigencia a Λ_1 consiste en la posibilidad de dividir las variables, entonces en calidad de Λ_1 se puede tomar un operador más general que el que hemos examinado. Si nos limitamos a las ecuaciones elípticas de segundo orden, entonces el caso más general de elección del operador Λ_1 corresponde a la aproximación de diferencias para el operador diferencial.

$$L_1 u = \frac{1}{k_2(x_1)} \frac{\partial}{\partial x_1} \left(k_1(x_1) \frac{\partial u}{\partial x_1} \right) + r(x_1) \frac{\partial u}{\partial x_1} - q(x_1) u,$$

cuyos coeficientes dependen solamente de x_1. Las condiciones de contorno sobre los lados $x_1 = 0$ y $x_1 = l_2$ del rectángulo $\overline{G}$ pueden ser cualquier combinación de condiciones de contorno de primero, segundo o tercer género (los coeficientes en la condición de contorno de tercer género deben ser constantes). Esto permite resolver problemas de contorno para la ecuación de Poisson en sistemas de coordenadas cilíndricos, esféricos, y polares.

§ 3. Método de reducción incompleta

1. Combinación de los métodos de Fourier y de reducción. El método de desarrollo en serie de aspecto singular construido en el punto 3 del § 2 nos permitió limitarnos solamente al cálculo de dos sumas de Fourier con un gasto de $O(N_1 N_2 \log_2 N_2)$ operaciones y a la resolución de una serie de problemas de contorno tripuntuales en $O(N_1, N_2)$ operaciones. Evidentemente, es posible un posterior perfecciona-

miento del método de separación de variables con vistas a disminuir el número de sumandos en las sumas a calcular y conservando la posibilidad de utilizar el algoritmo de la transformación de Fourier rápida.

Nosotros lograremos este objetivo combinando el método de desarrollo en serie de aspecto singular con el método de reducción estudiado en el capítulo III. Construyamos primeramente tal método combinado para el problema de Dirichlet más simple.

$$\begin{gathered} \Lambda u = -f(x), \quad x \in \omega, \quad u(x) = 0, \quad x \in \gamma, \\ \Lambda = \Lambda_1 + \Lambda_2, \quad \Lambda_\alpha u = u_{\bar{x}_\alpha x_\alpha}, \quad \alpha = 1, 2 \end{gathered} \tag{1}$$

sobre una red rectangular $\bar{\omega}$.

Para simplificar la discripción del método pasemos de la escritura puntual (escalar) del problema (1) a la vectorial.

Introduzcamos el vector de las incógnitas U_j de la siguiente forma: $U_j = (u(1, j), u(2, j), \ldots, u(N_1 - 1, j))$, $0 \leqslant j \leqslant N_2$, y definamos el vector F_j de los miembros derechos por la fórmula

$$F_j = (h_2^2 f(1, j), h_2^2 f(2, j), \ldots, h_2^2 f(N_1 - 1, j)),$$
$$1 \leqslant j \leqslant N_2 - 1.$$

Entonces el problema de diferencias (1) se puede escribir (véase el § 1 del cap. III) en forma del siguiente sistema de ecuaciones vectoriales:

$$\begin{gathered} -U_{j-1} + CU_j - U_{j+1} = F_j, \quad 1 \leqslant j \leqslant N_2 - 1, \\ U_0 = U_{N_2} = 0, \end{gathered} \tag{2}$$

donde la matriz tridiagonal cuadrática C se determina mediante las igualdades

$$CU_j = ((2E - h_2^2\Lambda_1)\, u(1, j), \ldots, (2E - h_2^2\Lambda_1)\, u(N_1 - 1, j)),$$
$$\Lambda_1 u = u_{\bar{x}_1 x_1}, \quad u(0, j) = u(N_1, j) = 0.$$

Sea N_2 múltiplo de 2: $N_2 = 2^m$. Recordemos, que el primer paso del proceso de exclusión en el método de reducción completa consiste (véase el § 2 del cap. III) en separar de (2) un sistema «reducido» para las incógnitas U_j con números j pares

$$-U_{j-2} + C^{(1)}U_j - U_{j+2} = F_j^{(1)}, \quad j = 2, 4, 6, \ldots, N_2-2$$
$$U_0 = U_{N_2} = 0$$

y las ecuaciones

$$CU_j = F_j + U_{j-1} + U_{j+1}, \quad j = 1, 3, 5, \ldots, N_2 - 1 \quad (4)$$

para determinar las incógnitas con números j impares. Aquí hemos designado

$$F_j^{(1)} = F_{j-1} + CF_j + F_{j+1}, \quad j = 2, 4, 5, \ldots, N_2 - 2, \quad (5)$$

$$C^{(1)} = [C]^2 - 2E. \quad (6)$$

Ocupémonos del sistema (3). Introduzcamos las notaciones

$$V_j = (v(1, j), v(2, j), \ldots, v(N_1 - 1, j)),$$

$$\Phi_j = (h_2^2\varphi(1, j), h_2^2 (2, j), \ldots, h_2^2 \varphi(N_1 - 1, j))$$

y pongamos

$$V_j = U_{2j}, \quad 0 \leqslant j \leqslant N_2/2, \quad \Phi_j = F_{2j}^{(1)}, \quad 1 \leqslant j \leqslant N_2/2 - 1,$$

$$v(0, j) = v(N_1, j) = 0, \quad 0 \leqslant j \leqslant N_2/2.$$

Estas notaciones permiten escribir el sistema (3) en la forma

$$-V_{j-1} + C^{(1)} V_j - V_{j+1} = \Phi_j, \quad j = 1, 2, \ldots, M_2 - 1$$

$$V_0 = V_M = 0, \quad (7)$$

donde $2M_2 = N_2$ y en virtud de (5)

$$\Phi_j = F_{2j-1} + CF_{2j} + F_{2j+1}, \quad j = 1, 2, \ldots, M_2 - 1. \quad (8)$$

Notemos ahora, que la función reticular $v(i, j)$ está definida para $0 \leqslant i \leqslant N_1$ y $0 \leqslant j \leqslant M_2$ y se anula para $j = 0$ y $j = M_2$. La función $\varphi(i, j)$ está definida para $1 \leqslant i \leqslant N_1 - 1$ y $1 \leqslant j \leqslant M_2 - 1$. Por eso estas funciones se pueden representar en forma de series de Fourier de aspecto singular

$$v(i, j) = \sum_{k_2=1}^{M_2-1} y_{k_2}(i)\, \mu_{k_2}^{(2)}(j), \quad 0 \leqslant i \leqslant N_1, \quad 0 \leqslant j \leqslant M_2,$$

$$\varphi(i, j) = \sum_{k_2=1}^{M_2-1} z_{k_2}(i)\, \mu_{k_2}^{(2)}(j), \quad (9)$$

$$1 \leqslant i \leqslant N_1 - 1, \quad 1 \leqslant j \leqslant M_2 - 1,$$

donde las funciones

$$\mu_{k_2}^{(2)}(j) = \frac{2}{\sqrt{l_2}} \operatorname{sen} \frac{k_2 \pi j}{M_2}, \quad k_2 = 1, 2, \ldots, M_2 - 1 \quad (10)$$

forman un sistema ortonormal en la red $\bar{\bar{\omega}}$ en el sentido del producto escalar

$$(u, v)=\sum_{j=1}^{M_2-1} u(j)\, v(j)\, h_2.$$

Los coeficientes de Fourier $z_{k_2}(i)$ de la función $\varphi(i, j)$ se encuentran por las fórmulas

$$\begin{gathered} z_{k_2}(i)=(\varphi, \mu_{k_2}^{(2)})=\sum_{j=1}^{M_2-1} h_2 \varphi(i, j)\, \mu_{k_2}^{(2)}(j), \\ 1 \leqslant k_2 \leqslant M_2-1, \quad 1 \leqslant i \leqslant N_1-1. \end{gathered} \tag{11}$$

De (9) obtenemos los siguientes desarrollos para los vectores V_j y Φ_j:

$$\begin{gathered} V_j=\sum_{k_2=1}^{M_2-1} Y_{k_2} \mu_{k_2}^{(2)}(j), \quad 0 \leqslant j \leqslant M_2, \\ \Phi_j=\sum_{k_2=1}^{M_2-1} h_2^2 Z_{k_2} \mu_{k_2}^{(2)}(j), \quad 1 \leqslant j \leqslant M_2-1, \end{gathered} \tag{12}$$

donde

$$Y_{k_2}=(y_{k_2}(1), \quad y_{k_2}(2), \ \ldots, \ y_{k_2}(N_1-1)),$$
$$Z_{k_2}=(z_{k_2}(1), \quad z_{k_2}(2), \ \ldots, \ z_{k_2}(N_2-1)).$$

Sustituyamos (12) en (7) y tengamos en cuenta las igualdades

$$\mu_{k_2}^{(2)}(j-1)+\mu_{k_2}^{(2)}(j+1)=2 \cos \frac{k_2 \pi}{M_2} \mu_{k_2}^{(2)}(j), \quad 1 \leqslant k_2 \leqslant M_2-1.$$

Obtenemos

$$\sum_{k_2=1}^{M_2-1}\left(C^{(1)}-2 \cos \frac{k_2 \pi}{M_2} E\right) Y_{k_2} \mu_{k_2}^{(2)}(j)=\sum_{k_2=1}^{M_2-1} h_2^2 Z_{k_2} \mu_{k_2}^{(2)}(j),$$

de donde, en virtud de la ortonormalidad del sistema (10) tendremos

$$\left(C^{(1)}-2 \cos \frac{k_2 \pi}{M_2} E\right) Y_{k_2}=h_2^2 Z_{k_2}, \quad 1 \leqslant k_2 \leqslant M_2-1. \tag{13}$$

Utilizando la relación (6) obtenemos

$$\begin{aligned} \left(C^{(1)}-2 \cos \frac{k_2 \pi}{M_2} E\right. & =[C]^2-2\left(1+\cos \frac{k_2 \pi}{M_2}\right) E= \\ & =\left(C-2 \cos \frac{k_2 \pi}{2 M_2} E\right)\left(C+2 \cos \frac{k_2 \pi}{2 M_2} E\right). \end{aligned}$$

Como la matriz $C^{(1)} - 2\cos\frac{k_2\pi}{M_2}E$ está factorizada, entonces para resolver la ecuación (13) se puede utilizar el algoritmo

$$\begin{aligned} &\left(C - 2\cos\frac{k_2\pi}{2M_2}E\right)W_{k_2} = h_2^2 Z_{k_2}, \\ &\left(C + 2\cos\frac{k_2\pi}{2M_2}E\right)Y_{k_2} = W_{k_2}, \quad 1 \leqslant k_2 \leqslant M_2 - 1, \end{aligned} \tag{14}$$

donde el vector auxiliar W_{k_2} tiene componentes $w_{k_2}(i)$:

$$W_{k_2} = (w_{k_2}(1), w_{k_2}(2), \ldots, w_{k_2}(N_1 - 1)),$$
$$w_{k_2}(0) = w_{k_2}(N_1) = 0.$$

Así han sido obtenidas las fórmulas necesarias. Pasando de la escritura vectorial a la escalar en (4), (8) y (14) y utilizando la relación $u(i, 2j) = v(i, j)$, que se deduce de la definición de V_j, obtendremos las siguientes fórmulas para calcular la función $\varphi(i, j)$ en el método construido:

$$\varphi(i, j) = f(i, 2j - 1) + 2f(i, 2j) + f(i, 2j + 1) - h_2^2\Lambda_1 f(1, 2j), \tag{15}$$

$$1 \leqslant j \leqslant N_2/2 - 1, \quad 1 \leqslant i \leqslant N_1 - 1, \quad f(0, 2j) = f(N_1, 2j) = 0,$$

las ecuaciones

$$\begin{aligned} &2\left(1 - \cos\frac{k_2\pi}{2M_2}\right)w_{k_2}(i) - h_2^2\Lambda_1 w_{k_2}(i) = h_2^2 z_{k_2}(i), \\ &1 \leqslant i \leqslant N_1 - 1, \\ &w_{k_2}(0) = w_{k_2}(N_1) = 0, \\ &2\left(1 + \cos\frac{k_2\pi}{2M_2}\right)y_{k_2}(i) - h_2^2\Lambda_1 y_{k_2}(i) = w_{k_2}(i), \\ &1 \leqslant i \leqslant N_1 - 1 \\ &y_{k_2}(0) = y_{k_2}(N_1) = 0 \end{aligned} \tag{16}$$

para determinar $y_{k_2}(i)$ con $k_2 = 1, 2, \ldots, M_2 - 1$ y las ecuaciones

$$2u(i, 2j - 1) - h_2^2\Lambda_1 u(i, 2j - 1) = h_2^2 f(i, 2j - 1) + u(i, 2j - 2) + u(i, 2j), \tag{17}$$
$$1 \leqslant i \leqslant N_1 - 1, \; u(0, 2j - 1) = u(N_1, 2j - 1) = 0$$

para encontrar la solución en $j = 1, 2, \ldots, M_2$. Para los coeficientes de Fourier $z_{k_2}(i)$ tenemos la fórmula (11) y de

(9) obtenemos

$$u(i, 2j) = \sum_{k_2=1}^{M_2-1} y_{k_2}(i)\, \mu_{k_2}^{(2)}(j), \quad 1 \leqslant j \leqslant M_2 - 1,$$

$$1 \leqslant i \leqslant N_1 - 1. \quad (18)$$

De esta manera, las fórmulas (10), (11), (15)—(18) describen completamente el método de resolución del problema (1), el cual es una combinación del método de desarrollo en serie de Fourier de aspecto singular y el método de reducción.

Pasemos ahora a la construcción del algoritmo del método. En las fórmulas (9), (16) y (18) hagamos el cambio $y_{k_2}(i) = a\bar{y}_{k_2}(i)$, $w_{k_2}(i) = a\bar{w}_{k_2}(i)$, $z_{k_2}(i) = a\bar{z}_{k_2}(i)$, donde $a = 2\sqrt{l_2}/N_2$, y en las fórmulas obtenidas omitamos el signo tilde. Este cambio permite librarse del factor normador $2/\sqrt{l_2}$, que se encuentra delante de la función propia $\mu_{k_2}^{(2)}(j)$ en las sumas (11) y (18). A continuación resolveremos los problemas (16) y (17) por el método de factorización. Es fácil cerciorarse de que aquí se cumplen las condiciones de corrección y estabilidad del método usual de factorización. Observemos una singularidad de los problemas (17). Ya que los coeficientes de la ecuación (17) no dependen de j, entonces los coeficientes de factorización α_i se deben calcular una vez al resolver el problema (17) para $j = 1$ y después utilizarlos en la resolución de las ecuaciones (17) para los j restantes.

Hagamos un resumen de las fórmulas de cálculo. Primeramente se calculan

$$\varphi(i, j) = f(i, 2j-1) + f(i, 2j+1) + 2\left(1 + \frac{h_2^2}{h_1^2}\right) f(i, 2j) -$$

$$- \frac{h_2^2}{h_1^2} [f(i-1, 2j) + f(i+1, 2j)],$$

$$1 \leqslant j \leqslant M_2 - 1, \quad 1 \leqslant i \leqslant N_1 - 1, \quad (19)$$

donde $f(0, 2j) = f(N_1, 2j) = 0$. Los valores $\varphi(i, j)$ se pueden situar en el lugar de $f(i, 2j)$. Las sumas

$$z_{k_2}(i) = \sum_{j=1}^{M_2-1} \varphi(i, j) \operatorname{sen} \frac{k_2 \pi j}{M_2}, \quad 1 \leqslant k_2 \leqslant M_2 - 1 \quad (20)$$

para $1 \leqslant i \leqslant N_1 - 1$ se calculan por el algoritmo de la transformación de Fourier discreta rápida, y $z_{k_1}(i)$ se coloca

en el lugar de $\varphi(i, k_2)$. Por el método de factorización

$$\alpha_{i+1} = 1/(c_{k_2} - \alpha_i), \quad \beta_{i+1} = [h_1^2 z_{k_2}(i) + \beta_i]\alpha_{i+1},$$
$$i = 1, 2, \ldots, N_1 - 1, \quad \alpha_1 = \beta_1 = 0,$$
$$w_{k_2}(i) = \alpha_{i+1} w_{k_2}(i+1) + \beta_{i+1}, \quad i = N_1 - 1, N_1 - 2, \ldots, 1, \qquad (21)$$
$$w_{k_2}(N_1) = 0, \quad c_{k_2} = 2 + 2\frac{h_1^2}{h_2^2} - 2\frac{h_1^2}{h_2^2}\cos\frac{k_2\pi}{N_2}$$

se resuelve la primera de las ecuaciones (16) y análogamente por las fórmulas

$$\alpha_{i+1} = \frac{1}{c_{k_2} - \alpha_i}, \quad \beta_{i+1} = \left[\frac{h_1^2}{h_2^2} w_{k_2}(i) + \beta_i\right]\alpha_{i+1},$$
$$i = 1, 2, \ldots, N_1 - 1, \quad \alpha_1 = \beta_1 = 0,$$
$$y_{k_2}(i) = \alpha_{i+1} y_{k_2}(i+1) + \beta_{i+1}, \quad i = N_1 - 1, N_2 - 1, \ldots, 1, \qquad (22)$$
$$y_{k_2}(N_1) = 0, \quad c_{k_2} = 2 + 2\frac{h_1^2}{h_2^2} + 2\frac{h_1^2}{h_2^2}\cos\frac{k_2\pi}{N_2}$$

se resuelve la segunda de las ecuaciones (16). Aquí los cálculos se realizan de una manera consecutiva para $k_2 = 1, 2, \ldots, M_2 - 1$ y los resultados $u_{k_2}(i)$ y $y_{k_2}(i)$ se colocan sucesivamente en el lugar de $z_{k_2}(i)$.

Para calcular las sumas

$$u(i, 2j) = \frac{4}{N_2}\sum_{k_2=1}^{M_2-1} y_{k_2}(i)\,\mathrm{sen}\,\frac{k_2\pi j}{M_2}, \quad 1 \leqslant j \leqslant M_2 - 1, \qquad (23)$$

con $1 \leqslant i \leqslant N_1 - 1$ de nuevo utilizamos el algoritmo de la transformación de Fourier rápida. Los problemas (17) se resuelven por el método de factorización teniendo en cuenta la singularidad señalada de estas ecuaciones:

$$\alpha_{i+1} = 1/(c - \alpha_i), \quad i = 1, 2, \ldots, N_1 - 1, \quad \alpha_1 = 0,$$
$$\beta_{i+1} = \left[h_1^2 f(i, 2j-1) + \frac{h_1^2}{h_2^2}(u(i, 2j-2) + u(i, 2j)) + \beta_i\right]\alpha_{i+1},$$
$$i = 1, 2, \ldots, N_1 - 1, \quad \beta_1 = 0. \qquad (24)$$
$$u(i, 2j-1) = \alpha_{i+1} u(i+1, 2j-1) + \beta_{i+1},$$
$$i = N_1 - 1, N_1 - 2, \ldots, 1, \quad u(N_1, 2j-1) = 0,$$
$$c = 2(1 + h_1^2/h_2^2)$$

para $1 \leqslant j \leqslant M_2$. La solución $u(i, j)$ se sitúa en el lugar de $f(i, j)$ y, por consiguiente, el algoritmo no exige memoria complementaria para la información intermedia.

Calculemos el número de operaciones aritméticas para el algoritmo (19)—(24). Para el cálculo según las fórmulas (19), (21), (22) y (24) se exigen $Q_{\pm} = (6{,}5\ N_2 - 9)\ (N_1 - 1)$ operaciones de suma y resta, $Q_* = (6\ N_2 - 8)\ (N_1 - 1)$ operaciones de multiplicación y $Q = (N_2 - 1)\ (N_1 - 1)$ operaciones de división. Para calcular las sumas (20) y (23) se exigen

$$Q_{\pm} = \left[\left(\frac{3}{2} \log_2 N_2 - \frac{7}{2} \right) N_2 - 2 \log_2 N_2 + 6 \right] (N_1 - 1)$$

operaciones de suma y resta y

$$Q_* = \left[\left(\frac{1}{2} \log_2 N_2 - 1 \right) N_2 + 1 \right] (N_1 - 1)$$

operaciones de multiplicación. En total el algoritmo (19)—(24) exige para $N_1 = N_2 = N = 2^n$

$$Q = (N^2 - 2N)(2 \log_2 N + 9) - 2N + 2 \log_2 N + 11$$

operaciones aritméticas. (25)

Para comparar citemos el número de operaciones del método de desarrollo en serie de aspecto singular (véase el punto 3 del § 2):

$$Q = \left(N^2 - \frac{3}{2} N \right) (4 \log_2 N + 2) - N + 2 \log_2 N + 2, \quad (26)$$

del método de desarrollo en serie doble (véase el punto 2 del § 2):

$$Q = \left(N^2 - \frac{3}{2} N \right) (8 \log_2 N - 10) + 5N + 4 \log_2 N - 10, \quad (27)$$

y también el número de operaciones para el segundo algoritmo del método de reducción completa (véase el cap. III, § 2, punto 4):

$$Q = \left(N^2 - \frac{11}{5} N \right) (5 \log_2 N + 5) + N + 6 \log_2 N + 5. \quad (28)$$

Si comparamos en las estimaciones (25)—(28) las constantes delante del término principal $N^2 \log_2 N$, entonces obtendremos, que el método combinado exige aproximadamente 4 veces menos operaciones aritméticas que el método de desarrollo en serie doble. Esta conclusión es cierta para los valores grandes de N. Para obtener relaciones reales entre los métodos examinados para los valores admisibles

de N presentemos la tabla 4 que contiene los valores de Q para estos métodos.

De esta forma, la combinación de los métodos de Fourier y de reducción permite disminuir el número de operaciones en comparación con el método inicial de desarrollo en serie de aspecto singular. Generalicemos este método combinado,

Tabla 4

Estimación / N	(25)	(26)	(27)	(28)
32	18 383	21 496	29 510	28 541
64	83 001	104 950	152 334	138 537
128	371 515	485 708	745 582	643 921

incluyendo en él l pasos de exclusión del método de reducción antes de desarrollar en serie de aspecto singular. Entonces el método expuesto en el punto 3 del § 2 se puede interpretar como un caso particular de este punto corresponde a $l = 0$, mientras que el método construido en este punto corresponde a $l = 1$. El método de reducción completa se puede examinar como el método para el cual $l = \log_2 N_2$.

Los datos de la tabla 4 muestran, que desde el punto de vista del gasto de operaciones aritméticas, existe un método generalizado óptimo para $1 \leqslant l < \log_2 N_2$. El análisis de las estimaciones para el número de operaciones en el método que contiene l pasos de reducción, dá el valor óptimo de $l = 1$ ó $l = 2$. Al mismo tiempo la ventaja insignificante en el número de operaciones del método para $l = 2$ puede perderse a causa de la creciente complejidad del algoritmo.

2. **Resolución de los problemas de contorno para la ecuación de Poisson en un rectángulo.** Examinemos ahora la aplicación del método construido en el punto 1 que permite solucionar los problemas de contorno para la ecuación de Poisson en un rectángulo. Supongamos que en la región $\overline{G} = \{0 \leqslant x_\alpha \leqslant l_\alpha,\ \alpha = 1, 2\}$ se exige hallar la solución de la ecuación

$$\frac{\partial^2 v}{\partial x_1^2} + \frac{\partial^2 v}{\partial x_2^2} = -\varphi(x), \quad x \in G, \tag{29}$$

que satisfaga las siguientes condiciones de contorno sobre la frontera Γ del rectángulo $\overline{G}$:

$$
\begin{aligned}
&\frac{\partial v}{\partial x_1} = \varkappa_{-1} v - g_{-1}(x_2), \quad x_1 = 0,\\
&-\frac{\partial v}{\partial x_1} = \varkappa_{+1} v - g_{+1}(x_2), \quad x_1 = l_1, \quad 0 \leqslant x_2 \leqslant l_2,\\
&\frac{\partial v}{\partial x_2} = -g_{-2}(x_1), \quad x_2 = 0, \qquad\qquad (30)\\
&-\frac{\partial v}{\partial x_2} = -g_{+2}(x_1), \quad x_2 = l_2, \quad 0 \leqslant x_1 \leqslant l_1,
\end{aligned}
$$

donde $\varkappa_{+1} \geqslant 0$, $\varkappa_{-1} \geqslant 0$, $\varkappa_{+1}^2 + \varkappa_{-1}^2 > 0$.

Supondremos, que en las condiciones (30) $\varkappa_{-1}$ y $\varkappa_{+1}$ son constantes. Bajo esta suposición las variables en el problema (29), (30) se separan.

Sobre la red rectangular $\overline{\omega} = \{x_{ij} = (ih_1, jh_2) \in \overline{G}, 0 \leqslant i \leqslant N_1, 0 \leqslant j \leqslant N_2, h_\alpha N_\alpha = l_\alpha, \alpha = 1, 2\}$ al problema (29)—(30) le corresponde el esquema de diferencias

$$\Lambda u \, (\Lambda_1 + \Lambda_2) u = -f(x), \quad x \in \overline{\omega}, \qquad (31)$$

donde $f(x) = \varphi(x) + \varphi_1(x) + \varphi_2(x)$,

$$
\Lambda_1 u = \begin{cases} \dfrac{2}{h_1}(u_{x_1} - \varkappa_{-1} u), & x_1 = 0,\\ u_{\bar{x}_1 x_1}, & h_1 \leqslant x_1 \leqslant l_1 - h_1,\\ \dfrac{2}{h_1}(-u_{\bar{x}_1} - \varkappa_{+1} u), & x_1 = l_1; \end{cases}
$$

$$
\Lambda_2 u = \begin{cases} \dfrac{2}{h_2} u_{x_2}, & x_2 = 0,\\ u_{\bar{x}_2 x_2}, & h_2 \leqslant x_2 \leqslant l_2 - h_2,\\ -\dfrac{2}{h_2} u_{\bar{x}_2}, & x_2 = l_2, \end{cases}
$$

y las funciones $\varphi_\alpha(x)$ se definen por las relaciones

$$
\varphi_\alpha(x) = \begin{cases} \dfrac{2}{h_\alpha} g_{-\alpha}(x_\beta), & x_\alpha = 0,\\ 0, \ h_\alpha \leqslant x_\alpha \leqslant l_\alpha - h_\alpha, & \beta = 3 - \alpha, \ \alpha = 1, 2,\\ \dfrac{2}{h_2} g_{+\alpha}(x_\beta), & x_\alpha = l_\alpha. \end{cases}
$$

En el capítulo III fue mostrado, que el esquema (31) en forma vectorial posee la siguiente escritura:

$$CU_0 = 2U_1 = F_0,$$
$$-U_{j-1} + CU_j - U_{j+1} = F_j, \quad 1 \leqslant j \leqslant N_2 - 1, \qquad (32)$$
$$-2U_{N_2-1} + CU_{N_2} = F_{N_2},$$

donde

$$U_j = (u(0, j), u(1, j), \ldots, u(N_1, j)),$$
$$F_j = (h_2^2 f(0, j), h_2^2 f(1, j), \ldots, h_2^2 f(N_1, j)),$$
$$CU_j = ((2E - h_2^2\Lambda_1)\, u(0, j), \ldots, (2E - h_2^2\Lambda_1) u(N_1, j)),$$
$$0 \leqslant j \leqslant N_2.$$

El sistema vectorial (32) se diferencia del sistema (2) examinado anteriormente por las condiciones de contorno y la definición de la matriz C. Sin embargo, construir un análogo del método del punto 1 para el problema (32) no presenta trabajo. Por cuanto la deducción de las fórmulas fundamentales para este método se diferencia solamente en detalles del citado en el punto 2, nosotros nos limitaremos a exponer un resumen de las fórmulas intermedias y definitivas principales. Para el método de reducción completa las fórmulas necesarias están descritas en el § 4 del cap. III.

Así, para los vectores $V_j = U_{2j}$, $0 \leqslant j \leqslant M_2$, donde $2M_2 = N_2$, después del paso de exclusión tendremos el problema

$$C^{(1)}V_0 - 2V_1 = \Phi_0,$$
$$-V_{j-1} + C^{(1)}V_j - V_{j+1} = \Phi_j, \quad 1 \leqslant j \leqslant M_2 - 1, \qquad (33)$$
$$-2V_{M_2-1} + C^{(1)}V_{M_2} = \Phi_{M_2},$$

donde el miembro derecho $\Phi_j = F_{2j}^{(1)}$, $0 \leqslant j \leqslant M_2$ se define por las fórmulas

$$\Phi_j = \begin{cases} CF_0 + 2F_1, & j = 0, \\ F_{2j-1} + CF_{2j} + F_{2j+1}, & 1 \leqslant j \leqslant M_2 - 1, \\ CF_{N_2} + 2F_{N_2-1}, & j = M_2. \end{cases}$$

Para los vectores V_j y Φ_j tenemos el desarrollo

$$V_j = \sum_{k_2=0}^{M_2} Y_{k_2} \mu_{k_2}^{(2)}(j), \quad \Phi_j = \sum_{k_2=0}^{M_2} h_2^2 Z_{k_2} \mu_{k_2}^{(2)}(j), \quad 0 \leqslant j \leqslant M_2,$$

dondo

$$\mu_{k_2}^{(2)}(j)=\begin{cases}\frac{2}{\sqrt{l_2}}\cos\frac{k_2\pi j}{M_2}, & 1\leqslant k_2\leqslant M_2-1,\\ \sqrt{\frac{1}{l_2}}\cos\frac{k_2\pi j}{M_2}, & k_2=0,\ M_2.\end{cases}$$

En virtud de (33) los coeficientes de Fourier de V_j y Φ_j están conectados por las relaciones

$$\left(C^{(1)}-2\cos\frac{k_2\pi}{M_2}E\right)Y_{k_2}=h_2^2Z_{k_2},\quad 0\leqslant k_2\leqslant M_2,$$

y las componentes del vector Z_{k_2} se expresan mediante las componentes del vector Φ_j de la siguiente forma:

$$z_{k_2}(i)=\sum_{j=1}^{M_2-1}h_2\varphi(i,j)\mu_{k_2}^{(2)}(j)+$$
$$+0{,}5h_2[\varphi(i,0)\mu_{k_2}^{(2)}(0)+\varphi(i,M_2)\mu_{k_2}^{(2)}(M_2)],\quad 0\leqslant i\leqslant N_1.$$

Al igual que antes, las incógnitas U_j con números j impares se determinan de las ecuaciones (4).

En las fórmulas obtenidas queda pasar a la escritura escalar y a la función propia no normada $\bar{\mu}_{k_2}^{(2)}(j)=\cos\frac{k_2\pi j}{M_2}$.

Como resultado obtendremos las siguientes fórmulas para el método de resolución del problema (31): para cada $0\leqslant i\leqslant$ $\leqslant N_1$ se calculan

$$\varphi(i,j)=\begin{cases}2[f(i,0)+f(i,1)]-h_2^2\Lambda_1f(i,0), & j=0,\\ f(i,2j-1)+f(i,2j+1)+2f(i,2j)-\\ -h_2^2\Lambda_1f(i,2j), & 1\leqslant j\leqslant M_2-1,\\ 2[f(i,N_2)+f(i,N_2-1)-h_2^2\Lambda_1f(i,N_2), & j=M_2,\end{cases}$$

se resuelven las ecuaciones

$$4\operatorname{sen}^2\frac{k_2\pi}{2N_2}w_{k_2}(i)-h_2^2\Lambda_1w_{k_2}(i)=h_2^2z_{k_2}(i),\quad 0\leqslant i\leqslant N_1,$$
$$4\cos^2\frac{k_2\pi}{2N_2}y_{k_2}(i)-h_2^2\Lambda_1y_{k_2}(i)=w_{k_2}(i),\quad 0\leqslant i\leqslant N_1$$

para $0\leqslant k_2\leqslant M_2$, donde

$$z_{k_2}(i)=\sum_{j=0}^{M_2}\rho_j\varphi(i,j)\cos\frac{k_2\pi j}{M_2},$$
$$0\leqslant k_2\leqslant M_2,\quad 0\leqslant i\leqslant N_1.$$

La solución $u(i, j)$ del problema (31) se determina por las fórmulas

$$u(i, 2j) = \sum_{k_2=0}^{M_2} \rho_{k_2} y_{k_2}(i) \cos\frac{k_2 \pi j}{M_2},$$

$$0 \leqslant j \leqslant M_2, \quad 0 \leqslant i \leqslant N_1$$

y de las ecuaciones

$$2u(i, 2j-1) - h_2^2 \Lambda_1 u(i, 2j-1) =$$

$$= h_2^2 f(i, 2j-1) + u(i, 2j-2) + u(i, 2j),$$

$$1 \leqslant j \leqslant M_2, \quad 0 \leqslant i \leqslant N_1.$$

Aquí se han utilizado las notaciones

$$\rho_j = \begin{cases} 1, & 1 \leqslant j \leqslant M_2 - 1, \\ 0{,}5, & j = 0,\ M_2,\ M_2 = 0{,}5 N_2, \end{cases}$$

y el operador Λ_1 está definido más arriba. Para encontrar $w_{k_2}(i)$, $y_{k_2}(i)$ y $u(i, 2j-1)$ aquí nosotros tenemos ecuaciones tripuntuales con las condiciones de contorno de tercer género, las cuales se resuelven por el método de factorización.

Notemos, que las fórmulas aquí citadas no cambian en absoluto si la red fuera no uniforme en la dirección de x_1. Cambiará solamente el tipo del operador Λ_1 y será un análogo en diferencias de la segunda derivada y de las condiciones de contorno de tercer género sobre una red no uniforme.

En general se debe observar, que se puede construir la correspondiente variante del método de separación de variables en todos los casos, excepto en uno, en las cuales se puede utilizar el método de reducción completa y con una estimación del número de operaciones $O(N^2 \log_2 N)$. La excepción la constituye el caso, en el cual por la dirección de exclusión de las incógnitas se dá una condición de contorno de tercer género al menos sobre uno de los lados del rectángulo.

3. Problema de Dirichlet de diferencias de alto orden de exactitud en un rectángulo. Examinemos otro ejemplo de aplicación del método de separación de variables. Supongamos que sobre la red rectangular $\bar{\omega}$ se exige hallar la solución *del problema de Dirichlet de elevado orden de exactitud para*

$$\Lambda u = \left(\Lambda_1 + \Lambda_2 + \frac{h_1^2+h_2^2}{12}\Lambda_1\Lambda_2\right) u = -f(u), \quad x \in \omega, \quad (34)$$

$$u(x) = 0, \quad x \in \gamma,$$

donde $\Lambda_\alpha u = u_{\bar{x}_\alpha x_\alpha}$, $\alpha = 1, 2$.
Para simplificar la condición de contorno está dada homogénea. El problema con condición de contorno no homogénea se reduce a (34) mediante la corrección del segundo miembro de la ecuación en los nodos fronterizos.

En el punto 4, § 1, cap. III fué obtenida la escritura vectorial del problema (34) en la siguiente forma:

$$-BU_{j-1} + AU_j - BU_{j+1} = F_j, \quad 1 \leqslant j \leqslant N_2 - 1,$$

$$U_0 = U_{N_2} = 0, \quad (35)$$

donde

$$U_j = (u(1, j), u(2, j), \ldots, u(N_1 - 1, j)), \; 0 \leqslant i \leqslant N_2,$$

$$F_j = (h_2^2 f(1, j), h_2^2 f(2, j), \ldots, h_2^2 f(N_1 - 1, j)), \; 1 \leqslant j \leqslant N_2 - 1$$

y las matrices B y A se definen por las relaciones

$$BU_j = \left(\left(E + \frac{h_1^2+h_2^2}{12}\Lambda_1\right) u(1, j), \ldots \right.$$
$$\left. \ldots, \left(E + \frac{h_1^2+h_2^2}{12}\Lambda_1\right) u(N_1 - 1, j)\right),$$

$$AU_j = \left(\left(2E - \frac{5h_2^2 - h_1^2}{6}\Lambda_1\right) u(1, j), \ldots \right.$$
$$\left. \ldots, \left(2E - \frac{5h_2^2 - h_1^2}{6}\Lambda_1\right) u(N_1 - 1, j)\right).$$

Las matrices A y B conmutan, es decir $AB = BA$.

Construyamos el método combinado de separación de variables para el problema (34). Primeramente realicemos el primer paso de exclusión del método de reducción para el sistema (35). Daremos la descripción de este paso independientemente de lo expuesto en el capítulo III. Escribamos tres ecuaciones consecutivas del sistema (35) para $j = 2, 4, 6, \ldots, N_2 - 2$:

$$-BU_{j-2} + AU_{j-1} - BU_j = F_{j-1},$$

$$-BU_{j-1} + AU_j - BU_{j+1} = F_j,$$

$$-BU_j + AU_{j+1} - BU_{j+2} = F_{j+1},$$

multipliquemos por B el primer miembro de la primera y la tercera ecuaciones, la del medio por A y sumémoslas. En virtud de la conmutabilidad de A y B, obtenemos

$$-B^2U_{j-1} + (A^2 - 2B^2)U_j - B^2U_{j+2} = F_j^{(1)},$$
$$j = 2, 4, 6, \ldots, N_2 - 2,$$
$$U_0 = U_{N_2} = 0,$$

donde $F_j^{(1)} = B(F_{j-1} + F_{j+1}) + AF_j$, $j = 2, 4, 6, \ldots$ $\ldots, N_2 - 2$. Designando, como es usual, $V_j = U_{2j}$, $0 \leqslant$ $\leqslant j \leqslant M_2$ y $\Phi_j = F_{2j}^{(1)}$, $1 \leqslant j \leqslant M_2 - 1$, donde $2M_2 = N_2$, escribamos este sistema en la forma

$$-B^2V_{j-1} + (A^2 - 2B^2)V_j - B^2V_{j+1} = \Phi_j, \ 1 \leqslant j \leqslant M_2 - 1,$$
$$V_0 = V_{M_2} = 0, \tag{36}$$

en este caso

$$\Phi_j = B *F_{2j-1} + F_{2j+1}) + AF_{2j}, \quad 1 \leqslant j \leqslant M_2 - 1. \tag{37}$$

Los restantes vectores desconocidos se encuentran de las ecuaciones

$$AU_{2j-1} = F_{2j-1} + B(U_{2j-2} + U_{2j}), \quad 1 \leqslant j \leqslant M_2. \tag{38}$$

Al igual que antes, resolveremos el sistema «reducido» (36) por el método de Fourier. Sustituyamos el desarrollo (12) en (36), donde las $\mu_{k_2}^{(2)}(j)$ están definidas en (10). Como resultado para los coeficientes de Fourier Y_{k_2} y Z_{k_2} de los vectores V_j y Φ_j obtendremos la relación

$$\left(A^2 - 4\cos^2\frac{k^2\pi}{2M_2}B^2\right)Y_{k_2} = h_2^2 Z_{k_2}, \quad 1 \leqslant k_2 \leqslant M_2 - 1 \tag{39}$$

que es el análogo de la relación (13), al mismo tiempo las componentes de los vectores Z_{k_2} y Φ_j están relacionados por la fórmula (11). Para resolver la ecuación (39) se puede utilizar el algoritmo

$$\begin{aligned} &\left(A - 2\cos\frac{k_2\pi}{2M_2}B\right)W_{k_2} = h_2^2 Z_{k_2}, \\ &\left(A + 2\cos\frac{k_2\pi}{2M_2}B\right)Y_{k_2} = W_{k_2}, \quad 1 \leqslant k \leqslant M_2 - 1. \end{aligned} \tag{40}$$

De esta manera, el método de resolución del problema (34) en forma vectorial se describe por las fórmulas (37), (11), (40), (12) y (38). Pasando a la escritura escalar y a la función propia no normada $\overline{\mu}_{k_2}^{(2)}(j) = \operatorname{sen}\frac{k_2\pi j}{M_2}$ por medio del

cambio del punto 1, obtenemos las siguientes fórmulas:

$$\varphi(i,\ j)=\left(E+\frac{h_1^2+h_2^2}{12}\Lambda_1\right)[f(i,\ 2j-1)+f(i,\ 2j+1)+$$
$$+2f(i,\ 2j)]-h_2^2\Lambda_1 f(i,\ 2j),\quad 1\leqslant j\leqslant M_2-1,$$
$$1\leqslant i\leqslant N_2-1,\qquad (41)$$
$$f(0,\ j)=0,\qquad 1\leqslant j\leqslant N_1-1$$

para el cálculo de $\varphi(i,\ j)$; las ecuaciones

$$4\operatorname{sen}^2\frac{k_2\pi}{2N_2}w_{k_2}(i)-h_2^2\left(1-\frac{4}{h_2^2}\operatorname{sen}^2\frac{k_2\pi}{2N_2}\times\right.$$
$$\left.\times\frac{h_1^2+h_2^2}{12}\right)\Lambda_2 w_{k_2}(i)=h_2^2 z_{k_2}(i),$$
$$1\leqslant i\leqslant N_1-1,\qquad w_{k_2}(0)=w_{k_2}(N_1)=0\qquad (42)$$

para el cálculo de $w_{k_2}(i)$ y

$$4\cos^2\frac{k_2\pi}{2N_2}y_{k_2}(i)-h_2^2\left(1-\frac{4}{h_2^2}\frac{k_2\pi}{2N_2}\frac{h_1^2+h_2^2}{12}\right)\Lambda_1 y_{k_2}(i)=$$
$$=w_{k_2}(i),$$
$$1\leqslant i\leqslant N_1-1,\qquad y_{k_2}(0)=y_{k_2}(N_2)=0\qquad (43)$$

para el cálculo de $y_{k_2}(i)$, las cuales se resuelven para $1\leqslant k_2\leqslant M_2-1$, donde

$$z_{k_2}(i)=\sum_{j=1}^{M_2-1}\varphi(i,\ j)\operatorname{sen}\frac{k_2\pi j}{M_2},\qquad 1\leqslant k_2\leqslant M_2-1,$$
$$1\leqslant i\leqslant N_1-1.\qquad (44)$$

La solución $u(i,\ j)$ del problema (34) se determina por las fórmulas

$$u(i,\ 2j)=\frac{4}{N_2}\sum_{k_2=1}^{M_2-1}y_{k_2}(i)\operatorname{sen}\frac{k_2\pi j}{M_2},$$
$$1\leqslant j\leqslant M_2-1,\qquad 1\leqslant i\leqslant N_1-1,\qquad (45)$$

y de las ecuaciones

$$2u(i,\ 2j-1)-\frac{5h_2^2-h_1^2}{6}\Lambda_1 u(i,\ 2j-1)=h_2^2 f(i,\ 2j-1)+$$
$$+\left(E+\frac{h_1^2+h_2^2}{12}\Lambda_1\right)[u(i,\ 2j-2)+u(i,\ 2j)],\qquad (46)$$
$$1\leqslant i\leqslant N_1-1,$$
$$u(0,\ 2j-1)=u(N_1,\ 2j-1)=0,\quad 1\leqslant j\leqslant M_2.$$

Nos queda por mostrar que las ecuaciones tripuntuales (42), (43) y (46) son solubles. Entonces para encontrar la solución se puede utilizar de método de factorización usual o el método de factorización no monótona.

Es suficiente mostrar, que para $1 \leqslant k_2 \leqslant N_2 - 1$ los valores propios del operador de diferencias

$$\mathcal{R} = \lambda_{k_2}^{(2)} E - \left(1 - \frac{h_1^2 + h_2^2}{12} \lambda_{k_2}^{(2)}\right) \Lambda_1,$$

$$\lambda_{k_2}^{(2)} = \frac{4}{h_2^2} \operatorname{sen}^2 \frac{k_2 \pi}{2N_2}$$

son diferentes de cero. En efecto, para $1 \leqslant k_2 \leqslant N_2/2 - 1$ el operador $h_2^2 \mathcal{R}$ coincide con el operador del problema (42), y para $k_2 = N_2/2$ coincide con el operador del problema (46). Si $N_2/2 + 1 \leqslant k_2 \leqslant N_2 - 1$, entonces el operador $h_2^2 \mathcal{R}$ tiene la forma

$$h_2^2 \mathcal{R} = 4 \operatorname{sen}^2 \frac{k_2 \pi}{2N_2} - h_2^2 \left(1 - \frac{h_1^2 + h_2^2}{12} \frac{4}{h_2^2} \operatorname{sen}^2 \frac{k_2 \pi}{2N_2}\right) \Lambda_1.$$

El cambio $k_2 = N_2 - k_2'$ da

$$h_2^2 \mathcal{R} = 2 \cos^2 \frac{k_2' \pi}{2N_2} - h_2^2 \left(1 - \frac{h_1^2 + h_2^2}{12} \frac{4}{h_2^2} \cos^2 \frac{k_2' \pi}{2N_2}\right) \Lambda_1,$$

donde $1 \leqslant k_2' \leqslant N_2/2 - 1$, es decir, en este caso el operador $h_2^2 \mathcal{R}$ coincide con el operador del problema (43).

Hallemos ahora los valores propios del operador $\mathcal{R}$ para un valor fijo k_2. Como los valores propios del operador Λ_1, para el caso de condiciones de contorno de primer género son (véase el § 5 del cap. I)

$$\lambda_{k_1}^{(1)} = \frac{4}{h_1^2} \operatorname{sen}^2 \frac{k_1 \pi}{2N_1}, \qquad k_1 = 1, 2, \ldots, N_1 - 1,$$

entonces los valores propios λ del operador $\mathcal{R}$ son

$$\lambda_{k_1 k_2} = \lambda_{k_1}^{(1)} + \lambda_{k_2}^{(2)} - \frac{h_1^2 + h_2^2}{12} \lambda_{k_1}^{(1)} \lambda_{k_2}^{(2)},$$

$$1 \leqslant k_1 \leqslant N_1 - 1, \qquad 1 \leqslant k_2 \leqslant N_2 - 1.$$

Ya que tienen lugar las siguientes estimaciones para los valores propios $\lambda_{k_1}^{(1)}$ y $\lambda_{k_2}^{(2)}$:

$$0 < \lambda_{k_\alpha}^{(\alpha)} < \frac{4}{h_\alpha^2}, \qquad \alpha = 1, 2,$$

entonces para todos k_1 y k_2 obtenemos fácilmente

$$\lambda_{k_1k_2}=\lambda_{k_1}^{(1)}\left(1-\frac{h_2^2}{12}\lambda_{k_2}^{(2)}\right)+\lambda_{k_2}^{(2)}\left(1-\frac{h_1^2}{12}\lambda_{k_1}^{(1)}\right)>$$
$$>2/3\,(\lambda_{k_1}^{(1)}+\lambda_{k_2}^{(2)})>0,$$

lo que se exigía demostrar.

Es fácil determinar, que para el problema (42) la condición suficiente de aplicabilidad del método de factorización usual posee la forma

$$1+\frac{2h_1^2-h_2^2}{3h_2^2}\operatorname{sen}^2\frac{k_2\pi}{2N_2}\geqslant 0$$

y, obviamente, está cumplido para cualquiera k_2. Para el problema (43) la condición análoga tiene la forma

$$1+\frac{2h_1^2-h_2^2}{3h_2^2}\cos^2\frac{k_2\pi}{2N_2}\geqslant 0$$

y lo mismo se cumple para todos los k_2. Al problema (46) le corresponde la condición (47) con $k_2 = 0{,}5\ N_2$. Por consiguiente, las problemas (42), (43) y (46) se pueden resolver por el método de factorización usual.

Capítulo V

Aparato matemático de la teoría de los métodos iterativos

El presente capítulo contiene los conocimientos y conceptos fundamentales de la teoría de los métodos iterativos que se exponen en los capítulos siguientes. En el § 1 son expuestos los conceptos más simples del análisis funcional, se exponen las propiedades fundamentales de los operadores lineales y no lineales en un espacio de Hilbert y también algunos teoremas sobre solubilidad de ecuaciones operacionales. En el § 2 se realiza un enfoque sistemático de los esquemas de diferencias como ecuaciones operacionales en un espacio de Hilbert abstracto y se indican las propiedades de los operadores correspondientes. En el § 3 se dan las definiciones y conceptos fundamentales de la teoría de los procesos iterativos, se examina la forma canónica de los esquemas iterativos y se dan los conceptos de convergencia y de número de iteraciones.

§ 1. Algunos conocimientos del análisis funcional

1. Espacios lineales. En los capítulos anteriores fueron estudiados los métodos directos fundamentales de resolución de las ecuaciones en diferencias más simples. Los métodos construidos se caracterizan, por el hecho de que con su ayuda es posible, en principio, obtener la solución exacta del problema de diferencias, realizando un número finito de operaciones. En este caso, naturalmente, se supone que la información entrante está dada con exactitud y todos los cálculos se llevan a efecto sin redondeo.

La efectividad de estos métodos es suficientemente alta, lo cual se alcanza teniendo en cuenta la estructura de la matriz del sistema a resolver. La exigencia de que se cumplan propiedades especiales de la matriz reduce el dominio

do aplicabilidad de estos métodos, limitándolo a los problemas más simples.

Para resolver problemas complejos y, en particular, problemas de diferencias no lineales, han obtenido una mayor difusión los métodos iterativos. La esencia de los métodos iterativos consiste en la construcción, por uno u otro método, de una sucesión de aproximaciones que converja a la solución, comenzando desde una cierta aproximación inicial. Para esto se toma como solución aproximada del problema la obtenida después de un número finito de iteraciones.

La universalidad de los métodos iterativos consiste ante todo, en que los últimos permiten resolver no solamente un problema concreto, sino una clase de problemas, que posean determinadas propiedades. Estas propiedades no se definen por la estructura de las ecuaciones reticulares, sino por las propiedades funcionales generales. Por cuanto en la mayoría de los métodos iterativos no se utiliza la estructura concreta de ecuaciones, entonces la teoría de los métodos iterativos se puede construir desde un punto de vista único, examinando en calidad de objeto inicial de investigaciones la ecuación operacional de primer género

$$Au = f,$$

donde A es un operador, f, un elemento prefijado y u, un elemento buscado de un cierto espacio H.

Antes de pasar a la construcción e investigación de los métodos iterativos daremos una breve lista de algunos aspectos del análisis funcional (sin demostraciones).

Se llama *espacio lineal* sobre el campo K de los números reales o complejos un conjunto H para cuyos elementos están definidas las operaciones, la suma de elementos y la multiplicación de un elemento por un número del campo K, cumpliéndose a su vez los siguientes axiomas (x, y, z son los elementos de H, λ y μ, los números de K):

1) ambas operaciones no hacen salir de H;
2) $x + y = y + x$, $x + (y + z) = (x + y) + z$ conmutatividad y asociatividad de la suma);
2) $\lambda(\mu x) = (\lambda\mu)x$ (asociatividad del producto);
4) $\lambda(x + y) = \lambda x + \lambda y$, $(\lambda + \mu)x = \lambda x + \mu x$, (distributividad del producto respecto de la suma);
5) existe un elemento 0 determinado unívocamente, tal que $x + 0 = x$ para todo $x \in H$;

6) para cada $x \in H$ existe un elemento $(-x) \in H$ unívocamente determinado, tal que $x + (-x) = 0$;
7) $1 \cdot x = x$.

En dependencia de por cuálos números, reales o complejos, se permite el producto de los elementos de H, obtenemos *un espacio lineal real o complejo.*

En los espacios lineales se puede introducir el concepto de dependencia o independencia lineal de sus elementos. Los elementos $x_1, x_2, \ldots, x_n$ de un espacio lineal H se llaman *linealmente independientes*, si de la igualdad:

$$\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n = 0 \qquad (1)$$

se deduce que $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$. Si, al contrario, se hallan $\lambda_1, \lambda_2, \ldots, \lambda_n$ no todos nulos, es decir, tiene lugar (1), entonces los elementos $x_1, x_2, \ldots, x_n$ se llaman *linealmente dependientes*.

El espacio H se llama *n-dimensional*, si en H existen n elementos linealmente independientes y cada $(n + 1)$-ésimo elemento es linealmente dependiente

Un conjunto cerrado no vacío H_1 de elementos de un espacio lineal H se llama *subespacio*, si el conjunto H_1 contiene junto con los elementos $x_1, x_2, \ldots, x_n$ cualquier combinación lineal $\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n$ de estos elementos.

La suma de un número finito de subespacios $H_1, H_2, \ldots, H_n$ es el conjunto de los elementos del tipo:

$$x = x_1 + x_2 + \ldots + x_n, \quad x_i \in H_i, \quad i = 1, 2, \ldots n. \qquad (2)$$

Dado que $H_1, H_2, \ldots, H_n$ son subespacios que pertenecen al espacio lineal H. Si cada elemento $x \in H$ se representa unívocamente en la forma (2), entonces se dice que H es *la suma directa de los subespacios $H_1, H_2, \ldots, H_n$* mientras que la expresión (2) se llama *desarrollo del elemento x en los elementos de $H_1, H_2, \ldots, H_n$*.

En este caso tendremos

$$H = H_1 \oplus H_2 \oplus \ldots \oplus H_n.$$

No es difícil mostrar, que si $H = H_1 \oplus H_2$, entonces H_1 y H_2 poseen únicamente el cero del espacio como único elemento común. Inversamente, si cualquier elemento $x \in H$ puede ser representado en la forma $x = x_1 + x_2$, $x_1 \in H_1$, $x_2 \in H_2$, y $H_1 \cap H_2 = 0$, entonces $H = H_1 \oplus H_2$.

El espacio lineal H se llama *normado*, si para cada elemento $x \in H$ está definido un número real $\| x \|$, llamado *norma*, el cual satisface las condiciones:

1) $\| x \| \geqslant 0$ y además $\| x \| = 0$, si $x = 0$;
2) $\| x + y \| \leqslant \| x \| + \| y \|$ (desigualdad triangular);
3) $\| \lambda x \| = | \lambda | \| x \|$, λ, un número.

La sucesión de elementos $\{x_n\}$ del espacio lineal normado H se llama *convergente* al elemento $x \in H$, si $\| x - x_n \| \to 0$ para $n \to \infty$. Si $\| x_n - x_m \| \to 0$ siendo n, $m \to \infty$, entonces la sucesión $\{x_n\}$ se llama *fundamental*.

Un espacio lineal normado H se llama *completo*, si cada sucesión fundamental $\{x_n\}$ de este espacio converge a un cierto elemento $x \in H$. Los espacios lineales normados completos se llaman espacios de *Banach*. Cada espacio lineal normado de dimensión finita es completo. Los subespacios de un espacio normado están normados de una manera natural.

Un mismo espacio lineal se puede normar por un conjunto infinito de procedimientos. Supongamos que en un espacio lineal se han introducido las normas $\| x \|_1$ y $\| x \|_2$ por dos métodos diferentes. Si existen constantes $0 < m \leqslant M$, tales que para cualquer $x \in H$ son ciertas las desigualdades

$$m \| x \|_1 \leqslant \| x \|_2 \leqslant M \| x \|_1,$$

entonces las normas se llaman *equivalentes*. Notemos que en un espacio de dimensión finita dos normas cualesquiera son equivalentes.

Si en un espacio lineal son introducidas dos normas equivalentes, entonces de la convergencia de cierta sucesión $\{x_n\}$ en una norma se deduce la convergencia en la otra norma.

Sea H un espacio lineal real (complejo) y supongamos que a cada dos elementos x, y de H se les hace corresponder un número real (complejo) (x, y), tal que:

1) $(x, y) = \overline{(y, x)}$ (simetría);
2) $(x + y, z) = (x, z) + (y, z)$ (distributividad);
3) $(\lambda x, y) = \lambda (x, y)$ (homogeneidad);
4) $(x, x) \geqslant 0$ para cualquier $x \in H$ y además $(x, x) = 0$, entonces y sólo entonces, cuando $x = 0$.

El número (x, y) se llama *producto escalar* de los elementos x e y. La raya encima significa el paso al número complejo conjugado.

Un espacio lineal normado H, en el cual la norma está generada por un producto escalar $\| x \| = \sqrt{(x, x)}$, se llama espacio *unitario* H. Un espacio unitario completo se llama

espacio de *Hilbert*. Un espacio unitario de dimensión finita es siempre completo.

Para el producto escalar es válida la desigualdad de Cauchy-Buniakovski $|(x, y)| \leqslant \|x\| \|y\|$. Los elementos x e y de un espacio unitario se llaman *mutuamente ortogonales*, si $(x, y) = 0$. El elemento $x \in H$ se llama *ortogonal al subespacio H_1 del espacio H*, si x es ortogonal a todo elemento $y \in H_1$. El conjunto H_2 de todos los elementos $x \in H$ ortogonales al subespacio H_1 del espacio H se llama *complemento ortogonal* del subespacio H_1. Notemos que el propio complemento ortogonal es un subespacio del espacio H.

Sea H_1 un subespacio arbitrario del espacio H, y H_2 el complemento ortogonal. Entonces H es la suma directa de H_1 y H_2, $H = H_1 \oplus H_2$. Por consiguiente, cada elemento $x \in H$ se representa de manera única en la forma $x = x_1 + x_2$, $x_\alpha \in H_\alpha$, $\alpha = 1, 2$ y al mismo tiempo $(x_1, x_2) = 0$.

El sistema $x_1, x_2, \ldots, x_n$, de elementos del espacio H se llama *sistema ortogonal*, si $(x_m, x_n) = \delta_{mn}$, $m, n, = 1, 2, \ldots$. donde δ_{mn} es el símbolo de Kronecker, igual a la unidad para $m = n$ y a cero para $m \neq n$.

Si no existe un elemento $x \in H$ diferente de cero y ortogonal a todos los elementos del sistema ortonormalizado $\{x_n\}$, entonces este sistema se llama *completo*. La serie de Fourier $\sum\limits_{k=1}^{\infty} c_k x_k$, donde $c_k = (x, x_k)$, $k = 1, 2, \ldots$, construida para cualquier $x \in H$ según el sistema completo ortonormalizado $\{x_n\}$, converge a este elemento, y para cualquier $x \in H$ tiene lugar la igualdad

$$\|x\|^2 = (x, x) = \sum_{k=1}^{\infty} c_k^2.$$

2. Operadores en espacios lineales normados. Sean X e Y espacios lineales normados. Se dice que sobre el conjunto $\mathscr{D} \subset X$ está prefijado un operador A con valores en Y (operador que actúa de $\mathscr{D}$ en Y), si a cada elemento $x \in \mathscr{D}$ se le pone en correspondencia un elemento $y = Ax \in Y$. El conjunto $\mathscr{D}$ se llama *dominio de definición del operador A* y se designa mediante $\mathscr{D}(A)$. El conjunto de todos los elementos $y \in Y$, representables en la forma $y = Ax$ $(x \in \mathscr{D}(A))$, se llama *campo de valores del operador A* y se designa por $\operatorname{im}(A)$. Si $\mathscr{D}(A) = X$ e $\operatorname{im}(A) \subset X$, es decir, si el operador A aplica X en sí mismo, entonces se dice que A es un operador en X. Si $\mathscr{D}(A) = X$ e $\operatorname{im}(A) = X$, es decir, si el ope-

rador A aplica X sobre sí mismo, entonces se dice que A es un operador sobre X.

El operador A se llama *lineal*, si $\mathscr{D}(A)$ es una variedad lineal en X y para cualesquiera $x_1, x_2 \in \mathscr{D}(A)$

$$A(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 A x_1 + \lambda_2 A x_2,$$

donde λ_1 y λ_2 son los números del campo K.

Un operador lineal A se llama *acotado* si existe una constante $M > 0$, tal que para cualesquiera $x \in \mathscr{D}(A)$

$$\| Ax \|_2 \leqslant M \| x \|_1, \tag{3}$$

donde $\| \cdot \|_1$ es la norma en X y $\| \cdot \|_2$, la norma en Y. Un operador no lineal A arbitrario se llama *acotado* sobre $\mathscr{D}(A)$, si

$$\sup_{x \in \mathscr{D}(A)} \| Ax \|_2 < \infty.$$

Para un operador lineal A la menor de las constantes M, que satisfacen la condición (3), se llama *norma* del operador y se designa por $\| A \|$. De la definición de la norma se deduce, que

$$\| A \| = \sup_{\| x \|_1 = 1} \| Ax \|_2 \quad \text{ó} \quad \| A \| = \sup_{x \neq 0} \frac{\| Ax \|_2}{\| x \|_1}.$$

Señalemos que en un espacio de dimensión finita todo operador lineal es acotado. Sea A un operador arbitrario que actúa de X en Y. El operador A se llama *continuo en el punto* $x \in X$, si de la condición $\| x_n - x \|_1 \to 0$ $(x_n \in X)$ se deduce que $\| Ax_n - Ax \|_2 \to 0$ para $n \to \infty$. Un operador lineal acotado es continuo.

Un operador arbitrario A satisface *la condición de Lipshitz con la constante* q, si

$$\| Ax_1 - Ax_2 \|_2 \leqslant q \| x_2 - x_2 \|_1, \quad x_1, x_2 \in \mathscr{D}(A). \tag{4}$$

Todo operador lineal acotado A satisface la condición de Lipschitz (4) si $q = \| A \|$.

Sea Λ un operador arbitrario, que actúa de X en Y. El operador lineal acotado $A'(x)$ se llama *derivada de Gato del operador* A *en el punto* x del espacio X, si para cualquier $z \in X$

$$\lim_{t \to 0} \left\| \frac{A(x + tz) - Ax}{t} - A'(x) z \right\|_2 = 0.$$

Con esto el campo de valores del operador $A'(x)$ pertenece a Y.

Si el operador A poseo derivada de Gato en cada punto del espacio X, entonces para cualesquiera $x_1, x_2 \in X$ es válida la desigualdad (4), donde $q = \sup\limits_{0 \leqslant t \leqslant l} \|A'(x_1 + t(x_2 - x_1))\|$ Si A es un operador lineal, entonces $A' = A$.

Todos los posibles operadores lineales acotados que actúan de X en Y, forman *un espacio lineal normado*, ya que la norma $\| A \|$ del operador A satisface todos los axiomas de la norma. Examinemos el conjunto de operadores lineales acotados que actúan de X en X. En este conjunto se puede introducir el producto AB de los operadores A y B del modo siguiente: $(AB)x = A(Bx)$. Es obvio que AB es un operador lineal acotado: $\| AB \| \leqslant \| A \| \, \| B \|$.

Si $(AB)x = (BA)x$ para todos los $x \in X$, los operadores A y B se llaman *permutables o conmutativos*; en este caso se escribe $AB = BA$.

Para resolver las ecuaciones del tipo $Ax = y$, se introduce el concepto de operador *inverso* A^{-1}. Sea A un operador de X en Y. Si a cada $y \in Y$ le corresponde únicamente un $x \in X$, para el cual $Ax = y$, entonces por esta correspondencia se define el operador A^{-1}, llamado *inverso* para A y que tiene el dominio de definición Y, y el campo de valores X.

Para cualesquiera $x \in X$, $y \in Y$ tenemos las identidades $A^{-1}(Ax) = x$, $A(A^{-1}y) = y$. No es difícil mostrar, que si A es lineal, entonces A^{-1} (si él existe) es también lineal.

LEMA 1. *Para que el operador A, el cual aplica X sobre Y, tenga el inverso, es necesario y suficiente, que $Ax = 0$ solamente para $x = 0$.*

TEOREMA 1. *Sea A un operador lineal de X en Y. Para que el operador inverso A^{-1} exista y sea acotado (como operador de Y en X), es menester y suficiente, que exista una constante $\delta > 0$, tal que*

$$\| Ax \|_2 \geqslant \delta \, \| x \|_1,$$

para todos los $x \in X$.

Además es válida la estimación $\|A^{-1}\| \leqslant 1/\delta$. Aquí $\| \cdot \|_1$ es la norma en X, y $\| \cdot \|_2$ es la norma en Y.

En otras palabras, para la existencia del operador inverso A^{-1} es necesario y suficiente, que la ecuación homogénea $Ax = 0$ posea sólo la solución trivial.

Sean A y B los operadores lineales acotados, que actúan en X y poseen inversos. Entonces $(AB)^{-1} = B^{-1}A^{-1}$.

Si el operador A es inversible, entonces tienen sentido las potencias A^k con exponentes enteros cualesquiera (y no

solamente no negativos). Precisamente, según la definición $A^{-k} = (A^{-1})^k$, $k = 1, 2, \ldots$. Las potencias de un mismo operador conmutan.

Introduzcamos el concepto de *núcleo* del operador lineal A. Se llama *núcleo del operador lineal* A el conjunto de todos los elementos x del espacio X, para los cuales $Ax = 0$. El núcleo del operador lineal A se designa con el símbolo ker A.

La condición ker $A = 0$ es menester y suficiente, para que el operador A posea el inverso.

El subespacio X_1 del espacio X se llama subespacio *invariante* del operador A, que actúa en X, si A no saca los elementos de X_1, es decir, $Ax \in X_1$, cuando $x \in X_1$.

Si el subespacio X_1 es invariante respecto al operador inversible A, entonces él es invariante respecto al operador A^{-1}.

Como ejemplos de subespacios invariantes del operador A pueden servir ker A o im A. Notemos, que si los operadores A y B conmutan, entonces los subespacios ker B e im B son invariantes con respecto al operador A.

El número

$$\rho(A) = \lim_{h\to\infty} \sqrt[h]{\| A^h \|}$$

se llama *radio espectral del operador lineal* A. El no depende de la definición de la norma y al mismo tiempo

$$\rho(A) = \inf_{\|\cdot\|} \| A \|.$$

Para todo operador lineal acotado A son válidas las desigualdades

$$\rho(A) \leqslant \| A \|, \quad \rho(A) \leqslant \sqrt[k]{\| A^k \|}, \quad k = 2, 3, \ldots.$$

LEMA 2. *Para que* $\| A \| = \rho(A)$, *es necesario y suficiente que* $\| A^k \| = \| A \|^k$, $k = 2, 3, \ldots$

Señalemos una propiedad más del radio espectral. Si los operadores A y B conmutan, entonces

$$\rho(AB) \leqslant \rho(A)\rho(B), \quad \rho(A+B) \leqslant \rho(A) + \rho(B).$$

3. Operadores en un espacio de Hilbert. Sea A un operador lineal acotado, que actúa en el espacio unitario H. De acuerdo con la definición general de la norma de un operador tenemos

$$\| A \| = \sup_{\|x\|=1} \| Ax \| = \sup_{x\in H} \sqrt{\frac{(Ax, Ax)}{(x, x)}}$$

y, por consiguiente, para cualquier $x \in H$, es válida la desigualdad

$$(Ax,\ Ax) \leqslant \| A \|^2 (x,\ x).$$

Utilizando la desigualdad de Cauchy-Buniakovski, obtenemos

$$|(Ax,\ x)| \leqslant \|Ax\| \|x\| \leqslant \| A \| (x,\ x). \tag{5}$$

A continuación *examinaremos solamente operadores acotados.*

El operador A^* se llama *conjugado* al operador A, si para cualesquiera $x,\ y \in H$ está cumplido la identidad

$$(Ax,\ y) = (x,\ A^*y).$$

Para todo operador lineal acotado A con dominio de definición $\mathscr{D}(A) = H$ existe un único operador A^* con dominio de definición $\mathscr{D}(A^*) = H$. El operador A^* es lineal y acotado, $\| A^* \| = \| A \|$.

Mostremos las propiedades fundamentales de la operación de conjugación: $(A^*)^* = A$, $(A + B)^* = A^* + B^*$, $(AB)^* = B^*A^*$, $(\lambda A)^* = \lambda A^*$. Si los operadores A y B conmutan, entonces conmutan los operadores conjugados A^* y B^*. Si A posee el inverso, entonces $(A^{-1})^* = (A^*)^{-1}$, es decir, las operaciones de tomar el inverso del operador y de conjugación son commutables.

LEMA 3. *Sea A un operador lineal en H. El espacio H puede ser representado en forma de sumas directas de los subespacios ortogonales*

$$H = \ker A \oplus \operatorname{im} A^*, \quad H = \ker A^* \oplus \operatorname{im} A.$$

En efecto, sea H_1 el complemento ortogonal de im A^* hasta del espacio H, es decir

$$H = H_1 \oplus \operatorname{im} A^*, \quad (x_1, x_2) = 0, \quad x_1 \in H_1, x_2 \in \operatorname{im} A^*.$$

Mostremos, que $H_1 = \ker A$. Sea $x_1 \in \ker A$, entonces para cualquier $x \in H$ tenemos $A^*x \in \operatorname{im} A^*$ y

$$(x_1,\ A^*x) = (Ax_1,\ x) = 0.$$

Por consiguiente, x_1 es ortogonal a im A^*, y por eso $x_1 \in H_1$. Por otra parte, sea $x_1 \in H_1$ (por lo tanto, x_1 es ortogonal a im A^*). Entonces para cualquier $x \in H$

$$0 = (x_1,\ A^*x) = (Ax_1,\ x).$$

Puesto que x es todo elemento arbitrario de H, entonces $Ax_1 = 0$ y, por consiguiente, $x_1 \in \ker A$. La primera afirmación del lema está demostrada. Análogamente se demuestra la segunda.

El operador lineal A se llama *autoconjugado* en H si $A = A^*$. Para un operador autoconjugado $(Ax, y) = (x, Ay)$ para todos los $x, y \in H$.

El operador A se llama *normal* si él conmuta con su conjugado, $A^*A = AA^*$, y *antisimétrico*, si $A^* = -A$. Los operadores autoconjugados y antisimétricos son normales.

Es conocido que si A y B son los operadores autoconjugados, entonces el operador AB es autoconjugado si y solamente si A y B son conmutables.

Si A es un operador lineal, entonces A^*A y AA^* son los operadores autoconjugados, además $\|A^*A\| = \|AA^*\| = \|A\|^2$ y

$$\ker A^*A = \ker A, \quad \operatorname{im} A^*A = \operatorname{im} A^*,$$
$$\ker AA^* = \ker A^*, \quad \operatorname{im} AA^* = \operatorname{im} A.$$

Todo operador A se puede representar en forma de la suma de un operador autoconjugado A_0 y otro antisimétrico A_1

$$A = A_0 + A_1,$$

donde $A_0 = 0{,}5\ (A + A^*)$, $A_1 = 0{,}5\ (A - A^*)$. Si H es un espacio real, entonces de aquí se deducen las igualdades

$$(Ax, x) = (A_0 x, x), \quad (A_1 x, x) = 0.$$

En un espacio complejo H tiene lugar la representación cartesiana del operador A:

$$A = A_0 + iA_1,$$

donde $A_0 = \operatorname{Re} A = \frac{1}{2}(A + A^*)$, $A_1 = \operatorname{Im} A = \frac{1}{2i}(A - A^*)$, son los operadores autoconjugados en H. Con esto son válidas las identidades:

$$\operatorname{Re}(Ax, x) = (A_0 x, x), \quad \operatorname{Im}(Ax, x) = (A_1 x, x),$$

para cualesquiera $x \in H$.

Si A es un operador autoconjugado en H, entonces tiene lugar la fórmula:

$$\|A\| = \sup_{x \neq 0} \frac{|(Ax, x)|}{(x, x)}, \quad x \in H.$$

LEMA 4. *Si A es un operador autoconjugado acotado en H, entonces para cualquier número entero $n > 0$, es válida la igualdad $\| A^n \| = \| A \|^n$.*

El lema 4 queda válido también para un operador normal.

De los lemas 2 y 4 se desprende, que para un operador normal (en particular, para un autoconjugado) A tiene lugar la igualdad $\rho(A) = \| A \|$.

LEMA 5. *Supongamos que en el espacio lineal H se ha introducido mediante dos procedimientos el producto escalar de los elementos x e y: $(x, y)_1$ y $(x, y)_2$. Si el operador A es autoconjugado en el sentido de cada producto escalar, entonces* $\| A \|_1 = \| A \|_2 = \rho(A)$.

El radio espectral da la estimación inferior a cualquier norma del operador. Introduzcamos el radio numérico del operador, el cual permite obtener estimaciones bilaterales para la norma.

El radio numérico del operador A, que actúa en el espacio complejo H, se define de la siguiente forma:

$$\bar{\rho}(A) = \sup_{\|x\|=1} |(Ax, x)|, \quad x \in H.$$

Para todo operador lineal acotado A son válidas las desigualdades:

$$\mu(A) \| A \| \leqslant \bar{\rho}(A) \leqslant \| A \|, \qquad \mu(A) \geqslant \frac{1}{2} \qquad y$$

además, $\bar{\rho}(A^n) \leqslant [\bar{\rho}(A)]^n$ para cualquier n natural. Si el operador A es autoconjugado, entonces $\bar{\rho}(A) = \| A \|$. Señalemos otra serie de propiedades interesantes del radio numérico. Así, por ejemplo, $\bar{\rho}(A^*) = \bar{\rho}(A)$, $\bar{\rho}(A^*A) = \| A \|^2$. Además, $\rho(A) \leqslant \bar{\rho}(A)$, donde $\rho(A)$ es el radio espectral del operador introducido anteriormente.

El operador lineal A, que actúa en el espacio de Hilbert H se llama *positivo* ($A > 0$), si $(Ax, x) > 0$ para todos los $x \in H$, excepto $x = 0$. En el caso de un espacio complejo H la definición de positividad se introduce sólo para los operadores autoconjugados, ya que de la positividad de un operador en este caso se deduce su autoconjugación.

Análogamente se introduce la definición del carácter *no negativo* del operador A (para todos los $x \in H$ $(Ax, x) \geqslant 0$) y del caracter *definido positivo* (para todos los $x \in H$ $(Ax, x) > \delta(x, x)$, donde $\delta > 0$). Un operador no lineal A que actua en H, se llama *monótono*, si

$$(Ax - Ay, x - y) \geqslant 0, \quad x, y \in H,$$

estrictamente monótono, si

$$(Ax - Ay, x - y) > 0, \quad x, y \in H, \quad x \neq y,$$

y *fuertemente monótono*, si para todos los $x, y \in H$ tiene lugar la desigualdad

$$(Ax - Ay, x - y) \geqslant \delta \| x - y \|^2, \quad \delta > 0.$$

TEOREMA 2. *Supongamos que el operador no lineal A posee derivada de Gato continua en cada punto $x \in H$. Entonces el operador A es fuertemente monótono sobre H, si y solamente si existe un $\delta > 0$, tal que*

$$(A'(x)y, y) \geqslant \delta (y, y), \quad y \in H.$$

Sea A un operador lineal no negativo. Al número (Ax, x) lo llamaremos *energía del operador*. Compararemos los operadores A y B por su energía. Si $((A - B)x, x) \geqslant 0$ para todos los $x \in H$, entonces escribiremos $A \geqslant B$.

Si existen constantes $\gamma_2 \geqslant \gamma_1 > 0$, tales que para los operadores lineales A y B son ciertas las desigualdades $\gamma_1 B \leqslant A \leqslant \gamma_2 B$, entonces a dichos operadores los llamaremos *energéticamente equivalentes* (en. eq.), y γ_1 y γ_2, constantes de equivalencia energética de los operadores A y B. Sean

$$\delta = \inf_{\|x\|=1} (Ax, x) \quad \text{y} \quad \Delta = \sup_{\|x\|=1} (Ax, x).$$

Los números δ y Δ se llaman *cotas* del operador A (autoconjugado en el caso de H complejo). Es evidente que son ciertas las desigualdades

$$\delta (x, x) \leqslant (Ax, x) \leqslant \Delta (x, x), \quad x \in H$$

ó

$$\delta E \leqslant A \leqslant \Delta E,$$

donde E es el operador identidad, $Ex = x$.

No es difícil cerciorarse de que la relación de desigualdad introducida en el conjunto de los operadores lineales que actúan en H, posee las siguientes propiedades:

1) de $A \geqslant B$ y $C \geqslant D$ se deduce $A + C \geqslant B + D$,
2) de $A \geqslant 0$ y $\lambda \geqslant 0$ se deduce $\lambda A \geqslant 0$,
3) de $A \geqslant B$ y $B \geqslant C$ se deduce $A \geqslant C$,
4) si $A > 0$ y A^{-1} existe, entonces $A^{-1} > 0$.

Además es obvio, que A^*A y AA^* son operadores no negativos para cualquier operador lineal A. Estos operadores serán positivos, si A es un operador positivo.

TEOREMA 3. *El producto AB de dos operadores A y B no negativos conmutables, uno de los cuales es autoconjugado, es también un operador no negativo.*

Para cualquier operador A autoconjugado no negativo tiene lugar una generalización de las desigualdad de Cauchy-Buniakovski

$$|(Ax, y)| \leqslant \sqrt{(Ax, x)}\sqrt{(Ay, y)}, \quad x, y \in H.$$

Sea D un operador autoconjugado positivo que actúa en H. Entonces se puede introducir *el espacio energético* H_D, que consiste de los elementos de H, con el producto escalar $(x, y)_D = (Dx, y)$ y la norma

$$\| x \|_D = \sqrt{(Dx, x)}.$$

Observemos, que si D es un operador autoconjugado, definido positivo y acotado en H, entonces para cualquier $x \in H$ en virtud de la desigualdad de Cauchy-Buniakovski son válidos las estimaciones

$$\delta (x, x) \leqslant (Dx, x) \leqslant \| Dx \| \; \| x \| \leqslant \Delta (x, x),$$
$$\Delta = \| D \|, \quad \delta > 0.$$

Estas desigualdades se pueden escribir en la forma

$$\sqrt{\delta} \| x \| \leqslant \| x \|_D \leqslant \sqrt{\Delta} \| x \|,$$

de donde se deduce que la norma usual $\| \cdot \|$ y la norma energética $\| \cdot \|_D$ son equivalentes.

Notemos, que se puede construir un espacio energético unitario H_D partiendo de un operador D positivo no autoconjugado. Para esto definiremos el producto escalar en H_D de la siguiente forma:

$$(x, y)_D = (D_0 x, y), \text{ donde } D_0 = 0{,}5\,(D + D^*).$$

Citemos una serie de lemas, los cuales contienen las desigualdades fundamentales, que nos serán necesarias en lo sucesivo.

LEMA 6. *Supongamos que para el operador lineal A está cumplido la condición $A \geqslant \delta E$, $\delta > 0$. Entonces para cualquier $x \in H$, tiene lugar la desigualdad*

$$(Ax, Ax) \leqslant \delta (Ax, x).$$

Si para un operador autoconjugado no negativo está cumplido la condición $A \leqslant \Delta E$, entonces para cualquier $x \in H$

tiene lugar la desigualdad

$$(Ax,\ Ax) \leqslant \Delta (Ax,\ x).$$

LEMA 7. *De la condición* $(Ax,\ Ax) \leqslant \Delta (Ax,\ x)$, $x \in H$, $\Delta > 0$ *para el operador no negativo* A se *deduce la desigualdad*

$$A \leqslant \Delta E,$$

y de la condición $(Ax,\ Ax) \geqslant \delta (Ax,\ x)$, $\delta > 0$, *para el operador autoconjugado no negativo* A *se deduce la desigualdad* $A \geqslant \delta E$.

COROLARIO 1. *De los lemas 6 y 7 se desprende, que para el operador* A *autoconjugado y definido positivo las desigualdades*

$$\delta E \leqslant A \leqslant \Delta E, \quad \delta > 0,$$

y

$$\delta (Ax,\ x) \leqslant (Ax,\ Ax) \leqslant \Delta (Ax,\ x), \quad \delta > 0,$$

son equivalentes.

COROLARIO 2. *De* (5) *y del lema* 6 *se deduce la estimación* $(Ax,\ Ax) \leqslant \| A \| (Ax,\ x)$, $x \in H$ *para el operador autoconjugado no negativo* A *en* H.

LEMA 8. *Sea* A *el operador positivo, acotado y autoconjugado en* H, $A > 0$, $\| Ax \| \leqslant \Delta \| x \|$. *Entonces el operador inverso* A^{-1} *es definido positivo* $A^{-1} \geqslant \frac{1}{\Delta} E$.

LEMA 9. *Sea* A *y* B *los operadores autoconjugados y definidos positivos en* H. *Entonces las desigualdades*

$$\gamma_1 B \leqslant A \leqslant \gamma_2 B, \quad \gamma_2 \geqslant \gamma_1 > 0$$

y

$$\gamma_1 A^{-1} \leqslant B^{-1} \leqslant \gamma_2 A^{-1}, \quad \gamma_2 \geqslant \gamma_1 > 0$$

son equivalentes.

LEMA 10 *Si* A *es el operador definido positivo* $A \geqslant \delta E$, $\delta > 0$, *entonces existe el operador inverso* A^{-1} y $\| A^{-1} \| \leqslant 1/\delta$. La demostración se deduce de la desigualdad

$$\delta \| x \|^2 \leqslant (Ax,\ x) \leqslant \| Ax \| \| x \|, \quad \delta > 0$$

y del teorema 1.

OBSERVACION. Si A es el operador positivo, entonces A^{-1} existe. En el caso de un espacio complejo H para la existencia del operador A^{-1} es suficiente la positividad de la componente real $A_0 = 0{,}5\ (A + A^*)$ o la positividad de la componente imaginaria $A_1 = \frac{1}{2i} (A - A^*)$ del operador A.

4. Funciones de un operador acotado. En la teoría de los métodos iterativos nos vemos obligados a tropezar con las funciones de un operador. Sea A el operador lineal acotado que actúa en el espacio normado X. Si $f(\lambda)$ es una función analítica entera de la variable λ, la cual se desarrolla en la serie $\sum_{k=0}^{\infty} a_k\lambda^k$, entonces se puede definir *la función* $f(A)$ *del operador* A con ayuda de la fórmula $f(A) = \sum_{k=0}^{\infty} a_k A^k$. El operador $f(A)$ también será lineal y acotado. En calidad de ejemplo citemos la función exponencial de un operador $e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$. La definición introducida de función de un operador se puede extender a una clase más amplia de funciones y trazar un cálculo operacional para operadores acotados. Nosotros daremos una definición más generalizada solamente para operadores acotados autoconjugados en el espacio de Hilbert.

Sean δ y Δ las cotas inferior y superior del operador A autoconjugado en H. Sea $f(\lambda)$ una función continua en el segmento $[\delta, \Delta]$. El operador $f(A)$ se llama *función del operador autoconjugado* A.

La correspondencia entre las funciones de la variable real y las funciones de un operador posee las siguientes propiedades:

1) Si $f(\lambda) = \alpha f_1(\lambda) + \beta f_2(\lambda)$, entonces $f(A) = \alpha f_1(A) + \beta f_2(A)$.

2) Si $f(\lambda) = f_1(\lambda) f_2(\lambda)$, entonces $f(A) = f_1(A) \times f_2(A)$.

3) De $AB = BA$ se deduce que $f(A)B = Bf(A)$ para cualquier operador lineal acotado B.

4) Si $f_1(\lambda) \leqslant f(\lambda) \leqslant f_2(\lambda)$ para todos los $\lambda \in [\delta, \Delta]$, entonces $f_1(A) \leqslant f(A) \leqslant f_2(A)$.

5) $\| f(A) \| \leqslant \max\limits_{\delta \leqslant \lambda \leqslant \Delta} | f(\lambda) |$.

6) $\bar{f}(A) = [f(A)]^*$, donde la raya encima de la función significa el paso a la función compleja conjugada. Si $f(\lambda)$ es una función real, entonces de aquí se deduce, que el operador $f(A)$ es autoconjugado en H.

De la propiedad 4) se deduce, que si $f(\lambda) \geqslant 0$ sobre $[\delta, \Delta]$, entonces $f(A)$ es el operador no negativo.

Un ejemplo importante de la función de un operador es

la raíz cuadrada del operador. El operador B se llama *raíz cuadrada del operador* A, si $B^2 = A$.

TEOREMA 4. *Existe una única raíz cuadrada autoconjugada no negativa de cualquier operador autoconjugado no negativo* A, *la cual conmuta con todo operador que conmute con* A.

Designaremos por $A^{\frac{1}{2}}$ la raíz cuadrada del operador A. Señalemos la siguiente propiedad:

$$\| A \| = \| A^{\frac{1}{2}} \|^2, \quad \text{si } A = A^* \geqslant 0.$$

TEOREMA 5 *Si* A *es el operador autoconjugado definido positivo,* $A = A^* \geqslant \delta E$, $\delta > 0$, *entonces existe el operador acotado autoconjugado* $A^{-\frac{1}{2}}$ $\| A^{-\frac{1}{2}} \| \leqslant 1/\sqrt{\delta}$.

La demostración se desprende de la desigualdad

$$\delta (x, x) \leqslant (Ax, x) = (A^{\frac{1}{2}} x, A^{\frac{1}{2}} x) = \| A^{\frac{1}{2}} x \|^2$$

y del teorema 1.

5. Operadores en un espacio de dimensión finita. Examinemos un espacio unitario H, n-dimensional. Supongamos que los elementos $x_1, x_2, \ldots, x_n$ forman una base ortonormal en H. Por la definición de espacio de dimensión finita cualquier elemento $x \in H$ se puede representar de manera única en forma de la combinación lineal

$$x = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n. \tag{6}$$

De la ortonormalidad del sistema $x_1, x_2, \ldots, x_n$ se deduce, que $c_h = (x, x_h)$.

De esta forma, a cada elemento $x \in H$ se le puede poner en correspondencia el vector $c = (c_1, c_2, \ldots, c_n)$, cuyas componentes son los coeficientes c_h del desarrollo (6).

Sea A el operador lineal, definido sobre H. En la base $x_1, x_2, \ldots, x_n$ a él le corresponde una matriz $\mathcal{A} = (a_{ik})$ de tamaño $n \times n$, donde $a_{ik} = (Ax_k, x_i)$. Inversamente toda matriz $\mathcal{A}$ de tamaño $n \times n$ define un operador lineal en H. En este caso al elemento Ax se le pone en correspondencia el vector

$$\left(\sum_{h=1}^{n} a_{1h} c_h, \sum_{l=1}^{n} a_{2h} c_h, \ldots, \sum_{l=1}^{n} a_{nh} c_h \right),$$

es decir, el vector $\mathcal{A}c$.

Si el operador A es autoconjugado en H, entonces la matriz $\mathcal{A}$ correspondiente es simétrica en cualquier base

ortonormal. Señalemos, que en una base no ortonormal a un operador autoconjugado A le corresponde una matriz no simétrica.

Detengámonos en las propiedades de los valores propios y elementos propios del operador lineal A. El número λ se llama *valor propio del operador* A, si la ecuación

$$Ax = \lambda x \qquad (7)$$

posee soluciones no nulas. El elemento $x \neq 0$, que satisface (7), se llama *elemento propio del operador* A, correspondiente al valor propio λ. Dicho de otra forma, los valores propios del operador A son aquellos valores λ, para los cuales ker $(A - \lambda E) \neq 0$; los elementos propios, correspondientes al valor propio λ, son los elementos diferentes de cero del subespacio ker $(A - \lambda E)$. Este mismo subespacio se llama *subespacio propio*, correspondiente al valor propio λ.

El conjunto $\sigma(A)$ de los valores propios del operador A se llama *espectro del operador* A.

1. Un operador autoconjugado A tiene n elementos propios ortonormalizados $x_1, x_2, \ldots, x_n$. Los valores propios correspondientes λ_k, $k = 1, 2, \ldots, n$ son reales. Si todos los valores propios son diferentes, entonces A se llama *operador con espectro simple*.

2. Para el operador autoconjugado A tienen lugar las igualdades

$$\|A\| = \rho(A) = \max_{1 \leqslant h \leqslant n} |\lambda_h|,$$

donde $\rho(A)$ es *el radio espectral del operador* A. Estas igualdades se conservan para un operador normal A.

3. Si $A = A^* \geqslant 0$, entonces todos los valores propios del operador A son no negativos. Con esto para cualquier $x \in H$

$$\delta(x, x) \leqslant (Ax, x) \leqslant \Delta(x, x),$$

donde $0 \leqslant \delta = \min_k \lambda_k$ y $\Delta = \max_h \lambda_h$.

Para el operador autoconjugado A se llama *relación de Relay* la expresión $(Ax, x)/(x, x)$.

Los valores propios, mayor y menor, del operador A se determinan con ayuda de la relación de Relay de la siguiente forma:

$$\delta = \min_{x \neq 0} \frac{(Ax, x)}{(x, x)}, \quad \Delta = \max_{x \neq 0} \frac{(Ax, x)}{(x, x)},$$

4. Designaremos mediante $\lambda(A)$ los valores propios del operador A. Sea $f(A)$ la función del operador autoconjugado A. Entonces $\lambda(f(A)) = f(\lambda(A))$ (teorema sobre la transformación de los espectros).

5. Si los operadores autoconjugados A y B conmutan, $A = A^*$, $B = B^*$ $AB = BA$, entonces ellos poseen un sistema común de elementos propios. En este caso los operadores AB y $A + B$ tienen el mismo sistema de elementos propios que los operadores A y B y los valores propios

$$\lambda(AB) = \lambda(A)\lambda(B), \quad \lambda(A + B) = \lambda(A) + \lambda(B).$$

6. Un elemento arbitrario $x \in H$ se puede desarrollar por elementos propios del operador autoconjugado A

$$x = \sum_{k=1}^{n} c_k x_k, \quad c_k = (x, x_k) \quad \text{y al mismo tiempo}$$

$$\|x\|^2 = \sum_{k=1}^{n} c_k^2.$$

El número λ se llama *valor propio del operador A con respesto al operador B*, si la ecuación

$$Ax = \lambda Bx \tag{8}$$

posee soluciones no nulas. El elemento $x \neq 0$ que satisface la ecuación (8), se llama *elemento propio del operador A con respecto al operador B* que corresponde al número λ.

7. Si los operadores A y B son autoconjugados en H, y el operador B, además, está definido positivamente, entonces existen n elementos propios $x_1, x_2, \ldots, x_n$ ortonormalizados en el espacio energético H_B: $(x_k, x_i)_B = \delta_{ki}$, $k, i = 1, 2, \ldots, n$. Los valores propios respectivos son reales y tienen lugar las desigualdades

$$\gamma_1 (Bx, x) \leqslant (Ax, x) \leqslant \gamma_2 (Bx, x),$$

donde

$$\gamma_1 = \min_k \lambda_k = \min_{x \neq 0} \frac{(Ax, x)}{(x, x)},$$

$$\gamma_2 = \max_k \lambda_k = \max_{x \neq 0} \frac{(Ax, x)}{(Bx, x)}.$$

Por consiguiente, las constantes de eq. en. de los operadores autoconjugados A y B en el caso del operador definido positivo B coinciden con los valores propios mínimo y máximo del problema generalizado (8).

6. Solubilidad de las ecuaciones operacionales. Supongamos que hace falta hallar la solución de la ecuación operacional de primer género

$$Au = f, \tag{9}$$

donde A es el operador lineal acotado en el espacio de Hilbert H, f es el elemento prefijado, y u, el elemento buscado de H. Supondremos que H es de dimensión finita. A nosotros nos interesará el problema sobre la solubilidad de la ecuación (9). Tiene lugar

TEOREMA 6. *Para que la ecuación* (9) *sea soluble para cualquier miembro derecho f, es necesario y suficiente, que la respectiva ecuación homogénea $Au = 0$, tenga únicamente la solución trivial $u = 0$. Con esto la solución de la ecuación* (9) *es única.*

La demostración del teorema se basa en el lema 1.

A la formulación del teorema se le puede dar otra forma: la ecuación (9) es soluble unívocamente para cualquier $f \in H$, entonces y solamente entonces, cuando $\ker A = 0$ (véase el punto 2).

Si $\ker A \neq 0$, entonces la ecuación es soluble sólo para restricciones complementarias sobre f. Recordemos, que en virtud del lema 3 el espacio H es la suma directa de subespacios ortogonales:

$$H = \ker A \oplus \operatorname{im} A^*, \quad H = \ker A^* \oplus \operatorname{im} A.$$

TEOREMA 7. *Para la solubilidad de la ecuación no homogénea* (9) *es necesario y suficiente que el miembro derecho f sea ortogonal al subespacio* $\ker A^*$. *En este caso la solución no es única y se determina con exactitud hasta un elemento arbitrario perteneciente al* $\ker A$:

$$u = \tilde{u} + \bar{u}, \quad \tilde{u} \in \ker A, \qquad A\bar{u} = f, \quad \bar{u} \in \operatorname{im} A^*.$$

Sea f ortogonal al $\ker A^*$. Se llama *solución normal de la ecuación* (9) la solución que posee norma mínima.

LEMA 11. *La solución normal es única y pertenece al subespacio* $\operatorname{im} A^*$ (*es decir, es ortogonal al* $\ker A$).

En efecto, sea $u = \tilde{u} + \bar{u}$, $\tilde{u} \in \ker A$, $\bar{u} \in \operatorname{im} A^*$. Entonces $\| u \|^2 = (u, u) = \| \tilde{u} \|^2 + \| \bar{u} \|^2 \geqslant \| \bar{u} \|^2$, ya que $\tilde{u}$ es un elemento arbitrario del subespacio $\ker A$. Por consiguiente, la norma $\| u \|$, será mínima si $u = \bar{u} \in \operatorname{im} A^*$.

Supongamos que no está cumplida la condición de ortogonalidad de f al subespacio $\ker A^*$. Entonces no existe la

solución de la ecuación (9) en el sentido clásico. Sea

$$f = \tilde{f} + \bar{f}, \quad \tilde{f} \in \ker A^*, \quad \bar{f} \in \operatorname{im} A.$$

Se llama *solución generalizada de la ecuación* (9), el elemento $u \in H$, para el cual $Au = \bar{f}$; la solución generalizada le concede el mínimo al funcional $\| Au - f \|$. En efecto, como $(Au - \bar{f}) \in \operatorname{im} A$ para cualquier $u \in H$, entonces

$$\| Au - f \|^2 = \| Au - \bar{f} \|^2 + \| \tilde{f} \|^2 \geqslant \| \tilde{f} \|^2,$$

y al mismo tiempo la desigualdad se alcanza si u es la solución generalizada.

La solución generalizada se determina con exactitud hasta un elemento arbitrario del subespacio $\ker A$. Llamaremos solución normal generalizada de la ecuación (9) la solución generalizada que posee norma mínima. La solución normal es única y pertenece a $\operatorname{im} A^*$.

Es evidente, que el concepto aquí introducido de solución normal, concuerda por completo con el dado más arriba. Señalemos, que si existe la solución normal clásica, entonces ella coincide con la solución normal generalizada.

Examinemos ahora la ecuación (9) con un operador no lineal arbitrario A, que actúa en el espacio de Hilbert H. En este caso, para demostrar la existencia y unicidad de la solución de la ecuación (9) con frecuencia, se utiliza *el principio de las aplicaciones contraídas de S. Banach.*

TEOREMA 8. *Sea un operador B definido en el espacio de Hilbert H y que aplica el conjunto cerrado T del espacio H en sí mismo. Supongamos, además, que el operador B es uniformemente contractante, es decir, satisface la condición de Lipschitz*

$$\| Bx - By \| \leqslant q \| x - y \|, \quad x, y \in T,$$

donde $q < 1$ y no depende de x e y. Entonces existe un y solamente un punto $x_* \in T$, tal que $x_* = Bx_*$. El punto x_* se llama punto fijo del operador B.

COROLARIO 1. *Si el operador B tiene derivada de Gato en H, la cual satisface la condición $\| B'(x) \| \leqslant q < 1$ para cualquier $x \in H$, entonces la ecuación $x = Bx$ tiene solución única en H.*

COROLARIO 2. *Supongamos que el operador C aplica el conjunto cerrado T en sí mismo y conmuta con el operador B, el cual satisface las condiciones del principio de las aplicaciones contraídas. Entonces el punto fijo del operador B es un punto*

fijo (posiblemente no único) del operador C. En particular, si alguna iteración B^n del operador B satisface el principio de las aplicaciones contraídas, entonces el punto fijo del operador B^n es también un punto fijo (único) del operador B.

Regresemos ahora a la resolución de la ecuación (9) con el operador no lineal A. Tiene lugar

TEOREMA 9. *Supongamos que el operador A posee derivada de Gato $A'(x)$ en cada punto $x \in H$ y existe $\tau \neq 0$, tal que se cumple la estimación $\| E - \tau A'(x) \| \leqslant q \leqslant 1$ para todos los $x \in H$. Entonces la ecuación (9) posee solución única en H.*

En efecto, la ecuación (9) se puede escribir en la siguiente forma:

$$u = u - \tau A u + \tau f, \quad \tau \neq 0. \tag{10}$$

Definamos el operador B: $Bx = x - \tau A x + \tau f$. Es evidente que el operador B posee derivada de Gato, igual a $B'(x) = E - \tau A'(x)$. En virtud de las condiciones del teorema tenemos $\|B'(x)\| \leqslant q < 1$ para cualquier $x \in H$. Por eso del corolario 1 del teorema 8 se desprende la existencia y unicidad de solución de la ecuación (10) y, por consiguiente, de la ecuación (9). El teorema está demostrado.

Señalemos, que en el capítulo VI serán examinados algunos métodos de obtención de estimaciones para las normas de operadores lineales del tipo $E - \tau C$, donde τ es un número.

En el principio de las aplicaciones contraídas no se agotan todos los casos, cuando existe solución de la ecuación no lineal. Al demostrar la solubilidad de la ecuación operacional (9) se puede utilizar una de las variantes del teorema sobre el punto fijo — *el principio de Brauder.*

TEOREMA 10. *Sea un operador B continuo monótono (estrictamente monótono) que satisface la condición*

$$(Bx, x) \geqslant 0 \text{ para } \| x \| = \rho > 0,$$

en el espacio de Hilbert de dimensión finita H. Entonces la ecuación $Bx = 0$ tiene al menos una (respectivamente única) solución en la bola $\| x \| \leqslant \rho$.

Utilicemos este teorema y formulemos las condiciones bajo las cuales la ecuación operacional (9) tiene solución única para cualquier miembro derecho f.

TEOREMA 11. *Sea dada la ecuación (9) con el operador continuo y fuertemente monótono A*

$$(Ax - Ay, x - y) \geqslant \delta \| x - y \|^2, \quad \delta > 0, \quad x, y \in H,$$

en el espacio de Hilbert de dimensión finita H. Entonces en la bola $\| u \| \leqslant \frac{1}{\delta} \| A0 - f \|$ *la ecuación (9) tiene solución única.*

En efecto, escribamos la ecuación (4) en la siguiente forma:

$$Bu = Au - f = 0.$$

Se ve, que el operador B es continuo y fuertemente monótono. Empleando la condición del teorema y la desigualdad de Cauchy-Buniakovski, obtenemos

$$(Bx, x) = (Ax - f, x) = (Ax - A0, x - 0) - (f - A0, x) \geqslant$$
$$\geqslant \delta \| x \|^2 - \| f - A0 \| \| x \| = (\delta \| x \| - \| A0 - f \|) \| x \|.$$

De aquí se deduce, que sobre la esfera $\| x \| = \frac{1}{\delta} \| A0 - f \|$ el operador B satisface la condición $(Bx, x) \geqslant 0$. Por eso en virtud del teorema 10 la ecuación $Bu = 0$ (y junto con ella la ecuación (9)) tiene solución única en la bola indicada. El teorema 11 está demostrado.

COROLARIO 1. *Si el operador A tiene derivada de Gato en H, que es el operador positivo definido en H, entonces las condicones del teorema 11 están cumplidas.*

En efecto, como en un espacio lineal de dimensión finita todo operador lineal es acotado, entonces la derivada de Gato es un operador continuo acotado y definido positivo en H. Del teorema 2 se desprende, que A es un operador fuertemente monótono. Además, de la acotación de la derivada de Gato se infiere, que el operador A satisface la condición de Lipschitz y por lo tanto es continuo.

§ 2. Esquemas de diferencias como ecuaciones operacionales

1. Ejemplos de espacios de funciones reticulares. En el § 1 del cap. I fueron introducidos los conceptos fundamentales de la teoría de los esquemas de diferencias: redes, ecuaciones reticulares, funciones reticulares, derivadas de diferencias, etc. La teoría forma los principios generales y reglas de construcción de esquemas de diferencias de una calidad dada. Un rasgo característico de esta teoría es la posibilidad de confrontar con cada ecuación diferencial toda una clase de esquemas de diferencias de las propiedades exigidas. Al construir la teoría general es natural librarse de la

estructura concreta y la forma explícita de las ecuaciones en diferencias. Esto conduce a la definición de los esquemas de diferencias como ecuaciones operacionales con operadores, que actúan en cierto espacio funcional, precisamente, en un espacio de funciones reticulares.

Por *espacio de funciones reticulares* se entiende el conjunto de las funciones definidas sobre una cierta red. Puesto que a cada función reticular se le puede poner en correspondencia el vector, cuyas coordenadas son los valores de la función reticular en los nodos de la red, entonces las operaciones de suma de funciones y de multiplicación de funciones por un número se definen de igual manera que para los vectores.

El espacio de las funciones reticulares es lineal, y si la red contiene un número finito de nodos, entonces el espacio es de dimensión finita. Su dimensión es igual al número de nodos de la red.

En el espacio de las funciones reticulares se puede introducir un producto escalar de funciones, convirtiendo este espacio en un espacio de Hilbert. Los distintos espacios de funciones reticulares se pueden diferenciar uno de otro por la elección de la red y la normación. Citemos algunos ejemplos.

EJEMPLO 1 Sea la red uniforme $\overline{\omega} = \{x_i = ih, 0 \leqslant i \leqslant N, hN = l\}$ con paso h introducida en el intervalo $0 \leqslant x \leqslant l$. Mediante ω, ω^+ y ω^- designaremos las siguientes partes de la red $\overline{\omega}$:

$$\omega = \{x_i \in \overline{\omega}, \quad 1 \leqslant i \leqslant N-1\},$$
$$\omega^+ = \{x_i \in \overline{\omega}, \quad 1 \leqslant i \leqslant N\},$$
$$\omega^- = \{x_i \in \overline{\omega}, \quad 0 \leqslant i \leqslant N-1\}.$$

En el conjunto H de las funciones reticulares, definidas sobre $\overline{\omega}$ y que toman valores reales, definiremos un producto escalar y una norma de la siguiente forma:

$$(u, v) = (u, v)_{\overline{\omega}} = \sum_{i=1}^{N-1} u_i v_i h + 0{,}5h\,(u_0 v_0 + u_N v_N),$$
$$\| u \| = \sqrt{(u, u)}, \quad u_i = u(x_i), \quad v_i = v(x_i).$$

Si u_i y v_i se consideran como los valores de las funciones $u(x)$ y $v(x)$ del argumento continuo $x \in [0, l]$ sobre la red $\overline{\omega}$, entonces el producto escalar (1) representa la fórmula de cuadratura de los trapecios para la integral $\int_0^l u(x)\,v(x)\,dx$.

Si las funciones reticulares están prefijadas sobre ω, ω^+, ω^-, entonces el producto escalar de funciones reticulares escalares se define respectivamente por las fórmulas

$$(u, v) = \sum_{i=1}^{N-1} u_i v_i h, \quad u, v \in H(\omega),$$

$$(u, v) = \sum_{i=1}^{N-1} u_i v_i h + 0{,}5 h u_N v_N, \quad u, v \in H(\omega^+),$$

$$(u, v) = \sum_{i=1}^{N-1} u_i v_i h + 0{,}5 h u_0 v_0, \quad u, v \in H(\omega^-),$$

Es fácil comprobar, que los productos escalares introducidos satisfacen todos los axiomas del producto escalar, y por lo tanto los espacios construidos son de Hilbert.

EJEMPLO 2. Sea ahora introducida en el segmento $0 \leqslant \leqslant x \leqslant l$ una red no uniforme arbitraria:

$$\bar{\omega} = \{x_i \in [0, l], \quad x_i = x_{i-1} + h_i, \quad 1 \leqslant i \leqslant N, \quad x_0 = 0, \quad x_N = l\}. \quad (2)$$

Recordemos la definición del paso medio $\hbar_i$ en el nodo x_i:

$$\hbar_i = 0{,}5\,(h_i + h_{i+1}), \quad 1 \leqslant i \leqslant N-1, \quad \hbar_0 = 0{,}5\,h_1, \quad \hbar_N = 0{,}5\,h_N. \quad (3)$$

Señalemos, que una red uniforme es un caso particular de la red no uniforme (2) para $h_i \equiv h$. Con esto tenemos $\hbar_i = h$, $1 \leqslant i \leqslant N-1$, $\hbar_0 = \hbar_N = 0{,}5h$.

Como más arriba, designemos mediante ω, ω^+, ω^- las partes respectivas de la red $\bar{\omega}$. Por analogía con el ejemplo 1 definiremos el producto escalar en los espacios reales de funciones reticulares dadas sobre las redes indicadas, por las fórmulas:

$$(u, v) = \sum_{i=0}^{N} u_i v_i \hbar_i, \quad u, v \in H(\bar{\omega}), \quad (4)$$

$$(u, v) = \sum_{i=1}^{N-1} u_i v_i \hbar_i, \quad u, v \in H(\omega), \quad (5)$$

$$(u, v) = \sum_{i=1}^{N} u_i v_i \hbar_i, \quad u, v \in H(\omega^+),$$

$$(u, v) = \sum_{i=0}^{N-1} u_i v_i \hbar_i, \quad u, v \in H(\omega^-).$$

Los espacios construidos de funciones reticulares son de *Hilbert* y tienen *dimensión finita*, igual al número de nodos de la red correspondiente.

Es cómodo escribir los productos escalares introducidos en la forma

$$(u, v) = \sum_{x_i \in \Omega} u(x_i)\, v(x_i)\, \hbar(x_i), \qquad u, v \in H(\Omega),$$

donde por Ω se entiende una de las redes $\bar{\omega}$, ω, ω^+ ó ω^-. Además de los productos escalares indicados con frecuencia se encuentran las sumas del tipo

$$(u, v)_{\omega^+} = \sum_{i=1}^{N} u_i v_i h_i, \qquad (u, v)_{\omega^-} = \sum_{i=0}^{N-1} u_i v_i h_{i+1}, \qquad (6)$$

las cuales se pueden utilizar en calidad de productos escalares en los espacios $H(\omega^+)$ y $H(\omega^-)$. Se ve, que para el producto escalar (4) en el espacio $H(\bar{\omega})$ es cierta la igualdad

$$(u, v) = 0{,}5\ [u\,(v)_{\omega^+} + (u, v)_{\omega^-}], \qquad u, v \in H(\bar{\omega}).$$

EJEMPLO 3 Sea una red rectangular no uniforme arbitraria $\bar{\omega} = \bar{\omega}_1 \times \bar{\omega}_2$ introducida en el rectángulo $\bar{G} = \{0 \leqslant x_\alpha \leqslant l_\alpha,\ \alpha = 1, 2\}$, donde

$$\bar{\omega}_\alpha = \{x_\alpha(i_\alpha) \in [0, l_\alpha],\quad x_\alpha(i_\alpha) = x_\alpha(i_\alpha - 1) + h_\alpha(i_\alpha),$$
$$1 \leqslant i_\alpha \leqslant N_\alpha,\quad x_\alpha(0) = 0,\quad x_\alpha(N_\alpha) = l_\alpha\},\ \alpha = 1, 2.$$

Sea $\hbar_\alpha(i_\alpha)$, $0 \leqslant i_\alpha \leqslant N_\alpha$ el paso medio en el nodo $x_\alpha(i_\alpha)$ por la dirección x_α:

$$\hbar_\alpha(i_\alpha) = 0{,}5\,[h_\alpha(i_\alpha) + h_\alpha(i_\alpha + 1)],\quad 1 \leqslant i_\alpha \leqslant N_\alpha - 1,$$
$$\hbar_\alpha(0) = 0{,}5\, h_\alpha(1),\quad \hbar_\alpha(N_\alpha) = 0{,}5\, \hbar_\alpha(N_\alpha),$$
$$\alpha = 1, 2.$$

En el espacio $H(\Omega)$ de las funciones reticulares, prefijadas sobre Ω, donde Ω es cualquier parte de la red $\bar{\omega}$, definiremos el producto escalar por la fórmula

$$(u, v) = \sum_{x_i \in \Omega} u(x_i)\, v(x_i)\, \hbar_1 \hbar_2, \qquad x_i = (x_1(i_1), x_2(i_2)).$$

En particular, si la red es uniforme en cada dirección, $h_\alpha(i_\alpha) \equiv h_\alpha$, $\alpha = 1, 2$, y si las funciones reticulares están definidas en ω (en los nodos interiores de la red $\bar{\omega}$), entonces

el producto escalar introducido se escribe en la forma

$$(u, v) = \sum_{i_1=1}^{N_1-1} \sum_{i_2=1}^{N_2-1} u(i_1, i_2)\, v(i_1, i_2)\, h_1 h_2, \quad u, v \in H(\omega).$$

Nosotros aquí nos limitaremos a los ejemplos citados. Otros ejemplos más complejos serán examinados en los capítulos siguientes al estudiar problemas de diferencias concretos.

2. Algunas identidades de diferencias. Pasemos ahora a la deducción de las fórmulas fundamentales, con ayuda de las cuales se transforman las expresiones que contienen funciones reticulares. Nosotros mostremos estas fórmulas para el caso, cuando las funciones reticulares están prefijadas sobre una red no uniforme, definida en (2).

Recordemos la definición de las derivadas fundamentales de diferencias de la función reticular:

$$y_{\bar{x},i} = \frac{y_i - y_{i-1}}{h_i}, \qquad y_{x,i} = y_{\bar{x},i+1} = \frac{y_{i+1} - y_i}{h_{i+1}},$$

$$y_{\check{x},i} = \frac{y_i - y_{i-1}}{\hbar_i}, \qquad y_{\hat{x},i} = \frac{y_{i+1} - y_i}{\hbar_i},$$

$$y_{\bar{x}\hat{x},i} = y_{x\check{x},i} = \frac{1}{\hbar_i}(y_{x,i} - y_{\bar{x},i}).$$

En el punto 2 del § 1 cap. I fueron obtenidas dos fórmulas de sumación por partes:

$$\sum_{i=m+1}^{n-1} u_{\hat{x},i} v_i \hbar_i = -\sum_{i=m+1}^{n} u_i v_{\bar{x},i} h_i + u_n v_n - u_{m+1} v_m, \tag{7}$$

$$\sum_{i=m+1}^{n-1} u_{\bar{x},i} v_i h_i = -\sum_{i=m}^{n-1} u_i v_{\hat{x},i} \hbar_i + u_{n-1} v_n - u_m v_m, \tag{8}$$

Sustituyendo en estas fórmulas las relaciones

$$h_i u_{\bar{x},i} = \hbar_i u_{\check{x},i}, \qquad \hbar_i u_{\hat{x},i} = h_{i+1} u_{x,i},$$

después de simples transformaciones obtenemos las fórmulas

$$\sum_{i=m+1}^{n-1} u_{\check{x},i} v_i \hbar_i = -\sum_{i=m}^{n-1} u_i v_{x,i} h_{i+1} + u_{n-1} v_n - u_m v_m, \tag{9}$$

$$\sum_{i=m+1}^{n-1} u_{x,i} v_i h_{i+1} = -\sum_{i=m+1}^{n} u_i v_{\check{x},i} \hbar_i + u_n v_n - u_{m+1} v_m, \tag{10}$$

$$\sum_{i=m+1}^{n} u_{\bar{x},i} v_i h_i = -\sum_{i=m}^{n-1} u_i v_{x,i} h_{i+1} + u_n v_n - u_m v_m. \tag{11}$$

Sustituyamos $m = 0$ y $n = N$ en las fórmulas (7), (9), (11) y tengamos en cuenta la definición (5) para el producto escalar en H (ω) y, además, la notación (6). Obtendremos las identidades

$$(u_x, v) = -(u, v_{\bar{x}})_{\omega+} + u_N v_N - u_1 v_0, \tag{7'}$$

$$(u_{\check{x}}, v) = -(u, v_x)_{\omega-} + u_{N-1} v_N - u_0 v_0, \tag{9'}$$

$$(u_{\bar{x}}, v)_{\omega+} = -(u, v_x)_{\omega-} + u_N v_N - u_0 v_0, \tag{11'}$$

para las funciones reticulares u_i y v_i, definidas sobre la red $\bar{\omega}$. Si en (7') ponemos $u_i = a_i y_{\bar{x},\,i}$ para $1 \leqslant i \leqslant N$, entonces obtendremos la primera fórmula de Green de diferencias

$$((a y_{\bar{x}})_{\hat{x}}, v) = -(a y_{\bar{x}}, v_{\bar{x}})_{\omega+} + a_N y_{\bar{x},\,N} v_N - a_1 y_{x,\,0} v_0. \tag{12}$$

Análogamente, suponiendo $u_i = a_i y_{x,\,i}$ en (9) para $0 \leqslant \leqslant i \leqslant N-1$, obtenemos

$$((a y_x)_{\check{x}}, v) = -(a y_x, v_x)_{\omega-} + a_{N-1} y_{\bar{x},\,N} v_N - a_0 y_{x,\,0} v_0.$$

Si de (12) restamos la igualdad

$$((y, (a v_{\bar{x}})_{\hat{x}}) = -(a y_{\bar{x}}, v_{\bar{x}})_{\omega+} + a_N v_{\bar{x},\,N} y_N - a_1 v_{x,\,0} y_0,$$

entonces obtendremos la segunda fórmula de Green de diferencias

$$((a y_{\bar{x}})_{\hat{x}}, v) - (y, (a v_{\bar{x}})_{\hat{x}}) = a_N (y_{\bar{x}} v - v_{\bar{x}} y)_N - a_1 (y_x v - v_x y)_0. \tag{13}$$

Señalemos, que para las funciones y_i y v_i, que se anulan para $i = 0$ e $i = N$ ($y_0 = y_N = 0$, $v_0 = v_N = 0$), la fórmula (12) tiene la forma

$$((a y_{\bar{x}})_{\hat{x}}, v) = -(a y_{\bar{x}}, v_{\bar{x}})_{\omega+},$$

mientras que la segunda fórmula de Green (13), tiene la forma

$$((a y_{\bar{x}})_{\hat{x}}, v) = (y, (a v_{\bar{x}})_{\hat{x}}).$$

En el caso general de las funciones reticulares arbitrarias, definidas sobre $\bar{\omega}$, las fórmulas (12) y (13) se pueden escribir en la forma

$$(\Lambda y, v) = -(a y_{\bar{x}}, v_{\bar{x}})_{\omega+}, \qquad \Lambda y, v) - (y, \Lambda y) = 0, \tag{14}$$

donde el operador de diferencias Λ, que aplica $H(\overline{\omega})$ sobre $H(\overline{\omega})$, se determina de la siguiente manera:

$$\Lambda(y_i) = \begin{cases} \frac{1}{\hbar_0} a_1 y_{x,0}, & i=0, \\ (a y_{\bar{x}})_{\hat{x},i}, & 1 \leqslant i \leqslant N-1, \\ -\frac{1}{\hbar_N} a_N y_{\bar{x},N}, & i=N. \end{cases}$$

Aquí el producto escalar en $H(\overline{\omega})$ está prefijado por la fórmula (4). Observemos que la igualdad (14) expresa la autoconjugación del operador Λ en el espacio $H(\overline{\omega})$.

Hemos examinado el caso, cuando sobre la red las funciones reticulares toman valores reales. Si ellas adquieren valores complejos sobre $\overline{\omega}$, entonces se introduce el espacio de Hilbert complejo $H(\overline{\omega})$ con el producto escalar

$$(u, v) = \sum_{i=0}^{N} u_i \bar{v}_i \hbar_i, \quad u, v \in H(\overline{\omega}), \tag{15}$$

donde $\bar{v}_i$ es el número complejo conjugado de v_i. Análogamente se define el producto escalar en $H(\omega)$,

$$(u, v) = \sum_{i=1}^{N-1} u_i \bar{v}_i \hbar_i, \quad u, v \in H(\omega), \tag{16}$$

y también en $H(\omega^+)$ y $H(\omega^-)$. Con esto las fórmulas de sumación por partes (7'), (9') y (11') toman la forma

$$(u_{\hat{x}}, v) = -(u, v_{\bar{x}})_{\omega^+} + u_N \bar{v}_N - u_1 \bar{v}_0,$$

$$(u_{\check{x}}, v) = -(u, v_x)_{\omega^-} + u_{N-1} \bar{v}_N - u_0 \bar{v}_0,$$

$$(u_{\bar{x}}, v)_{\omega^+} = -(u, v_x)_{\omega^-} + u_N \bar{v}_N - u_0 \bar{v}_0,$$

y las fórmulas de Green de diferencias toman la forma:

$$((a y_{\bar{x}})_{\hat{x}}, v) = -(a y_{\bar{x}}, v_{\bar{x}})_{\omega^+} + a_N y_{\bar{x},N} \bar{v}_N - a_1 y_{x,0} \bar{v}_0$$

$$((a y_{\bar{x}})_{\hat{x}}, v) - (y, (a v_{\bar{x}})_{\hat{x}}) = ((\bar{a} - a) y_{\bar{x}}, v_{\bar{x}})_{\omega^+} +$$
$$+ (a y_{\bar{x}} \bar{v} - \bar{a} y \bar{v}_{\bar{x}})_N - (a_1 y_{x,0} \bar{v}_0 - \bar{a}_1 y_0 \bar{v}_{x,0}).$$

Aquí se ha utilizado la notación (16).

Utilizando el operador Λ introducido anteriormente y la notación (15) para el producto escalar en $H(\overline{\omega})$, la segunda fórmula de Green

de diferencias se puede escribir en la forma

$$(\Lambda y,\ v)-(y,\ \Lambda v)=((\bar{a}-a)\ y_{\bar{x}},\ v_{\bar{x}})_{\omega+}.$$

De aquí se deduce, que en el espacio de Hilbert complejo $H\ (\overline{\omega})$ el operador Λ es autoconjugado si todos los a_i son reales.

Las relaciones análogas a la primera y segunda fórmulas de Green de diferencias (12) y (13), tienen lugar también para el operador de diferencias $(ay_{\bar{x}\hat{x}})_{\bar{x}\hat{x}}$. Mostremos, por ejemplo, el análogo de la fórmula (12)

$$\sum_{i=2}^{N-2}(ay_{\bar{x}\hat{x}})_{\bar{x}\hat{x},\ i}v_i\hbar_i=\sum_{i=1}^{N-1}a_iy_{\bar{x}\hat{x},\ i}v_{\bar{x}\hat{x},\ i}\hbar_i+$$

$$+[(ay_{\bar{x}\hat{x}})_{\bar{x}}v-ay_{\bar{x}\hat{x}}v_x]_{N-1}-[(ay_{\bar{x}\hat{x}})_x v-ay_{\bar{x}\hat{x}}v_{\bar{x}}]_1.$$

3. Cotas de los operadores de diferencias más simples. Al estudiar las propiedades de los operadores de diferencias se necesitan desigualdades, que dan las estimaciones para las cotas de los operadores y para las constantes de equivalencia energética de dos operadores que actúan en el espacio H de funciones reticulares.

Examinemos primeramente los operadores de diferencias dados sobre un conjunto de funciones reticulares de un argumento y definidos en la red uniforme $\overline{\omega}=\{x_i=ih\in[0,\ l],\ 0\leqslant i\leqslant N,\ hN=l\}$. Más adelante se utilizarán las notaciones

$$(u,\ v)=\sum_{i=1}^{N-1}u_iv_ih+0{,}5h\,(u_0v_0+u_Nv_N),$$

$$(u,\ v)_{\omega+}=\sum_{i=1}^{N}u_iv_ih.$$

Tiene lugar

LEMA 12. *Para toda función $y_i=y\ (x_i)$, definida sobre la red uniforme $\overline{\omega}$ y que se anula para $i=0$, e $i=N$ son válidas las desigualdades*

$$\gamma_1(y,\ y)\leqslant(y_{\bar{x}}^2,\ 1)_{\omega+}\leqslant\gamma_2(y,\ y),\tag{17}$$

donde

$$\gamma_1=\frac{4}{h^2}\,\text{sen}^2\frac{\pi}{2N}\geqslant\frac{8}{l^2},\qquad \gamma_2=\frac{4}{h^2}\cos^2\frac{\pi}{2N}<\frac{4}{h^2}.$$

En efecto, sea $\mu_k(i)$ la función propia ortonormalizada del problema

$$(\mu_k)_{\bar{x}x} + \lambda_k\mu_k = 0, \qquad 1 \leqslant i \leqslant N-1, \tag{18}$$
$$\mu_k(0) = \mu_k(N) = 0.$$

En el punto 1 del § 5 del cap. I fue subrayado, que la función y_i que satisface las condiciones del lema, puede ser representada en forma de la suma

$$y_i = \sum_{k=1}^{N-1} c_k\mu_k(i), \qquad c_k = (y, \mu_k). \tag{19}$$

De (18) y (19) hallamos

$$y_{\bar{x}x,\,i} = \sum_{k=1}^{N-1} c_k(\mu_k)_{\bar{x}x,\,i} = -\sum_{k=1}^{N-1} \lambda_k c_k\mu_k(i), \qquad 1 \leqslant i \leqslant N-1.$$

Utilizando la ortonormalidad de las funciones propias μ_k, obtenemos

$$(y, y) = \sum_{k=1}^{N-1} c_k^2, \qquad -(y_{\bar{x}x}, y) = \sum_{k=1}^{N-1} \lambda_k c_k^2. \tag{20}$$

En virtud de la primera fórmula de Green de diferencias (12) tendremos

$$-(y_{\bar{x}x}, y) = (y_{\bar{x}}^2, 1)_{\omega^+}. \tag{21}$$

Los valores propios λ_k del problema (18) fueron hallados en el punto 1 del § 5 del cap. I:

$$\lambda_k = \frac{4}{h^2}\operatorname{sen}^2\frac{k\pi h}{2l} = \frac{4}{h^2}\operatorname{sen}^2\frac{k\pi}{2N}, \qquad 1 \leqslant k \leqslant N-1,$$

y al mismo tiempo

$$\gamma_1 = \min_k \lambda_k = \lambda_1 = \frac{4}{h^2}\operatorname{sen}^2\frac{\pi}{2N},$$

$$\gamma_2 = \max_k \lambda_k = \lambda_{N-1} = \frac{4}{h^2}\cos^2\frac{\pi}{2N}.$$

De aquí y de (20) y (21) se deducen las estimaciones (17) del lema 12.

OBSERVACIÓN 1. Las estimaciones (17) son exactas en el sentido, que ellas pasan a ser igualdades, si en lugar de y_i se toman $\mu_1(i)$ y $\mu_{N-1}(i)$. Señalemos que $\gamma_1 = 8/l^2$, si $h = l/2$, es decir, para $N = 2$. Para $N = 4$ tenemos $\gamma_1 = 32/(l^2(2+\sqrt{2})) > 8/l^2$.

OBSERVACIÓN 2. Si y_i se anula únicamente cuando $i = 0$ ó $i = N$, entonces en (17) tenemos

$$\gamma_1 = \frac{4}{h^2}\,\mathrm{sen}^2\frac{\pi}{4N} \geqslant \frac{8}{l^2(2+\sqrt{2})},$$

$$\gamma_2 = \frac{4}{h^2}\cos^2\frac{\pi}{4N} < \frac{4}{h^2}.$$

Si y_i es la función reticular arbitraria sobre $\bar{\omega}$, entonces en (17) tenemos $\gamma_1 = 0$ y $\gamma_2 = 4/h^2$. Para la demostración de estas afirmaciones se debe examinar en lugar del problema (18) el problema correspondiente de valores propios, estudiado en el § 5 del cap. I.

Las desigualdades (17) se pueden escribir en la forma

$$\gamma_1(y, y) \leqslant (-\Lambda y, y) \leqslant \gamma_2(y, y), \tag{22}$$

si se introduce el operador Λ por la fórmula $\Lambda y_i = y_{\bar{x}x,\,i}$, $1 \leqslant i \leqslant N-1$, sobre las funciones y_i, que satisfacen las condiciones $y_0 = y_N = 0$. Si la función reticular y_i se anula solamente en un extremo de la red $\bar{\omega}$, entonces el operador Λ se debe definir por las fórmulas

$$\Lambda y_i = \begin{cases} y_{\bar{x}x,\,i}, & 1 \leqslant i \leqslant N-1, \\ -\frac{2}{h}\,y_{\bar{x},\,i}, & i = N, \text{ si } y_0 = 0, \end{cases} \tag{23}$$

o

$$\Lambda y_i = \begin{cases} \frac{2}{h}\,y_{x,\,i}, & i = 0, \\ y_{\bar{x}x,\,i}, & 1 \leqslant i \leqslant N-1, \text{ si } y_N = 0, \end{cases}$$

Teniendo en cuenta, que en cada uno de estos casos de la primera fórmula de Green de diferencias se desprenden las igualdades $(\Lambda y, y) = (y_{\bar{x}}^2, 1)_{\omega^+}$, obtenemos las desigualdades (22), donde γ_1 y γ_2 están indicadas en la observación 2, e y_i se anula en el extremo correspondiente de la red $\bar{\omega}$.

Si y_i es la función reticular arbitraria, entonces el operador Λ se conviene definir así:

$$\Lambda y_i = \begin{cases} \frac{2}{h}\,y_{x,\,0}, & i = 0, \\ y_{\bar{x}x,\,i}, & 1 \leqslant i \leqslant N-1, \\ -\frac{2}{h}\,y_{\bar{x},\,N}, & i = N. \end{cases}$$

En este caso también son ciertas las desigualdades (22) y

$$(-\Lambda y,\ y) = -(y_{\bar{x}x},\ y) + y_{\bar{x},N}\, y_N - y_{x,0}\, y_0 = (y_{\bar{x}}^2,\ 1)_{\omega^+}.$$

Las constantes γ_1 y γ_2 están indicadas en la observación 2.

Así, hemos hallado las cotas para los operadores de diferencias más simples. Mostremos ahora que para todos los operadores introducidos en este punto es válida la desigualdad

$$|(-\Lambda u,\ v)| \leqslant (-\Lambda u,\ u)^{\frac{1}{2}} (-\Lambda v,\ v)^{\frac{1}{2}}. \qquad (24)$$

Ilustraremos la idea de obtener la desigualdad (24) en el ejemplo del operador $\Lambda y = y_{\bar{x}x}$. Introduzcamos el espacio $H(\omega)$ de las funciones reticulares definidas sobre ω con el producto escalar $(u,\ v) = \sum\limits_{i=1}^{N-1} u_i v_i h$, $u,\ v \in H(\omega)$. Al operador de diferencias Λ en el espacio $H(\omega)$ le corresponde el operador lineal A, definido por la igualdad

$$A y_i = -\Lambda \mathring{y}_i, \quad 1 \leqslant i \leqslant N-1,$$

donde $y \in H(\omega)$, $y_i = \mathring{y}_i$ para $1 \leqslant i \leqslant N-1$, y $\mathring{y}_0 = \mathring{y}_N = 0$. El operador A aplica $H(\omega)$ sobre $H(\omega)$.

En virtud de la igualdad $(u,\ v) = (\mathring{u},\ \mathring{v})$, tenemos $(Au,\ v) = -(\Lambda\mathring{u},\ \mathring{v})$, donde $\mathring{u}_0 = \mathring{u}_N = 0$ y $\mathring{v}_0 = \mathring{v}_N = 0$. De (22) se deduce, que $(Au,\ u) \geqslant \gamma_1 (u,\ u)$, $\gamma_1 > 0$. De esta forma, el operador A es definido positivo en $H(\omega)$.

Demostremos, que él es autoconjugado en $H(\omega)$. En efecto, de la segunda fórmula de Green de diferencias (13) tendremos

$$(Au,\ v) = -(\Lambda\mathring{u},\ \mathring{v}) = -(\mathring{u}_{\bar{x}x},\ \mathring{v}) = -(\mathring{u},\ \mathring{v}_{\bar{x}x}) = (u,\ Av).$$

Ya que para un operador autoconjugado no negativo es válida la desigualdad generalizada de Cauchy-Buniakovski $|(Au,\ v)| \leqslant (Au,\ u)^{\frac{1}{2}} (Av,\ v)^{\frac{1}{2}}$, entonces de aquí obtendremos

$$|(-\Lambda\mathring{u},\ \mathring{v})| \leqslant (-\Lambda\mathring{u},\ \mathring{u})^{\frac{1}{2}} (-\Lambda\mathring{v},\ \mathring{v})^{\frac{1}{2}},$$

lo cual era preciso demostrar.

4. Estimaciones inferiores para algunos operadores de diferencias. En el lema 12 de hecho se han hallado las constantes de equivalencia energética del operador unidad E y el operador A, el cual corresponde al operador de diferencias $-\Lambda y = -y_{\bar{x}x}$ sobre las funciones que se anulan en los extremos de la red ω, es decir, γ_1 y γ_2 de las desigualdades

$$\gamma_1 E \leqslant A \leqslant \gamma_2 E.$$

Ahora obtendremos desigualdades, que relacionan los operadores A y D donde $Dy_i = \rho_i y_i$, $1 \leqslant i \leqslant N-1$ y $\rho_i \geqslant 0$. Para eso necesitamos definir la función de Green de diferencias del operador Λ.

Supongamos que, sobre la red $\bar{\omega}$ introducida más arriba, se exige hallar la solución del problema de diferencias

$$\Lambda v_i = v_{\bar{x}x,\, i} = -f_i, \quad 1 \leqslant i \leqslant N-1,$$
$$v_0 = v_N = 0. \tag{25}$$

La función reticular G_{ik}, que para $k = 1, 2, \ldots, N-1$ fijo, satisface las condiciones

$$\Lambda G_{ik} = G_{\bar{x}x,\, ik} = -\frac{1}{h}\delta_{ik}, \quad 1 \leqslant i \leqslant N-1,$$
$$G_{0k} = G_{Nk} = 0,$$

donde δ_{ik} es el símbolo de Kronecker:

$$\delta_{ik} = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases}$$

se llama *función de Green del operador de diferencias* Λ.

Exponemos las propiedades fundamentales de la función de Green:

1) la función de Green es simétrica, $G_{ik} = G_{ki}$ y, además, G_{ik} como función de k para $i = 1, 2, \ldots, N-1$ fija satisface las condiciones

$$\Lambda G_{ik} = G_{\bar{x}x,\, ik} = -\frac{1}{h}\delta_{ik}, \quad 1 \leqslant k \leqslant N-1$$
$$G_{i0} = G_{iN} = 0.$$

2) la función de Green es positiva, $G_{ik} > 0$ para $i, k \neq 0, N$.

3) para cualquier función reticular que satisfaga las ondiciones $y_0 = y_N = 0$, es cierta la representación

$$y_i = -\sum_{k=1}^{N-1} G_{ik} \Lambda y_k h, \tag{26}$$

de manera que la solución del problema (25) es representable en la forma

$$v_i = \sum_{k=1}^{N-1} G_{ik} f_k h, \quad 0 \leqslant i \leqslant N.$$

Esta afirmación se demuestra con ayuda de la segunda fórmula de Green de diferencias (13) y la propiedad 1).

LEMA 13 *Sea $\rho_i \geqslant 0$ la función reticular, definida sobre ω y no idénticamente igual a cero. Para toda función y_i, definida sobre $\bar{\omega}$ y que satisfaga las condiciones $y_0 = y_N = 0$, es cierta la estimación*

$$\gamma_1 (\rho y, y) \leqslant (y_{\bar{x}}^2, 1)_{\omega^+}, \tag{27}$$

donde $1/\gamma_1 = \max\limits_{1 \leqslant i \leqslant N-1} v_i$, y v_i *es la solución del problema de contorno*

$$\begin{aligned} &\Lambda v_i = v_{\bar{x}x, i} = -\rho_i, \quad 1 \leqslant i \leqslant N-1, \\ &v_0 = v_N = 0. \end{aligned} \tag{28}$$

En efecto, sea $y_0 = y_N = 0$. Empleando (26), obtenemos

$$(\rho y, y) = \sum_{i=1}^{N-1} \rho_i y_i^2 h = -\sum_{i=1}^{N-1} \rho_i y_i h \left(\sum_{k=1}^{N-1} G_{ik} \Lambda y_k h\right) =$$
$$= -\sum_{k=1}^{N-1} h \Lambda y_k \left(\sum_{i=1}^{N-1} \rho_i y_i G_{ik} h\right) = -(\Lambda y, w),$$

donde hemos designado $w_k = \sum\limits_{i=1}^{N-1} \rho_i y_i G_{ik} h, \quad 0 \leqslant k \leqslant N.$ Aplicando la desigualdad (24), de aquí halleremos

$$(\rho y, y) \leqslant (-\Lambda y, y)^{\frac{1}{2}} (-\Lambda w, w)^{\frac{1}{2}},$$

o en virtud de (21)

$$(\rho y, y)^2 \leqslant (y_{\bar{x}}^2, 1)_{\omega^+} (-\Lambda w, w). \tag{29}$$

Utilicemos la propiedad 1) de la función de Green G_{ik}. Obtenemos

$$-\Lambda w_k = -\sum_{i=1}^{N-1} h \rho_i y_i \Lambda G_{ik} = \sum_{i=1}^{N-1} \rho_i y_i \delta_{ik} = \rho_k y_k$$

y por consiguiente,

$$(-\Lambda w,\ w)=\sum_{k=1}^{N-1} h\rho_k y_k\Big(\sum_{i=1}^{N-1} h\rho_i y_i G_{ik}\Big)-\sum_{i=1}^{N-1}\sum_{k=1}^{N-1} a_{ik}y_i y_k,$$

donde hemos designado $a_{ik}=h^2\rho_i\rho_k G_{ik}$, $1\leqslant i,\ k\leqslant N-1$. Mediante la desigualdad: $2y_iy_k\leqslant y_i^2+y_k^2$, y también la simetría y la positividad de la función de Green G_{ik}, hallamos

$$(-\Lambda w,\ w)\leqslant\sum_{i=1}^{N-1} 0{,}5y_i^2\sum_{k=1}^{N-1} a_{ik}+\sum_{k=1}^{N-1} 0{,}5y_k^2\sum_{i=1}^{N-1} a_{ki}=$$

$$=\sum_{i=1}^{N-1} y_i^2\sum_{k=1}^{N-1} a_{ik}=\sum_{i=1}^{N-1}\rho_i y_i^2 h\Big(\sum_{k=1}^{N-1}\rho_k G_{ik}h\Big).$$

En virtud de la propiedad 3) la solución del problema (28) se escribe en la forma

$$v_i=\sum_{k=1}^{N-1}\rho_k G_{ik}h>0,\quad 1\leqslant i\leqslant N-1.$$

Por lo tanto,

$$(-\Lambda w,\ w)=\sum_{i=1}^{N-1}\rho_i y_i^2 v_i h\leqslant\max_{1\leqslant i\leqslant N-1} v_i\,(\rho y,\ y)=\frac{1}{\gamma_1}(\rho y,\ y).$$

De aquí y de (29) se desprende la estimación (27) del lema.

OBSERVACIÓN 1. Se puede mostrar, que la función $v_i=0{,}5x_i\,(l-x_i)$, donde $x_i=ih\in[0,\ l]$, es la solución del problema (28) para $\rho_i\equiv 1$. De aquí se deduce la estimación

$$\gamma_1(y,\ y)\leqslant(y_{\bar x}^2,\ 1)_{\omega^+},\quad \gamma_1=8/l^2,\quad y_0=y_N=0. \qquad (30)$$

OBSERVACIÓN 2. El lema 13 se generaliza para el caso, cuando y_i se anula sólo en un extremo de la red $\bar\omega$. Por ejemplo, si $y_0=0$, entonces en (27) tenemos $1/\gamma_1=\max_{1\leqslant i\leqslant N} v_i$, donde v_i es la solución del problema $\Lambda v_i=-\rho_i$, $1\leqslant i\leqslant N$, $v_0=0$ con el operador de diferencias Λ, definido en (23).

LEMA 14. *Sean $\rho_i\geqslant 0$, $d_i\geqslant 0$ definidas sobre ω, y la función $a_i\geqslant c_1>0$ definida sobre ω^+. Para toda función y_i, definida sobre $\bar\omega$ y que satisfaga las condiciones $y_0=$*

$= y_N = 0$, *es cierta la estimación*

$$\gamma_1(\rho y,\ y) \leqslant (ay_{\bar{x}}^2,\ 1)_{\omega+} + (dy,\ y),\quad 1/\gamma_1 = \max_{1\leqslant i\leqslant N-1} v_i,$$

donde v_i es la solución del problema de contorno

$$\Lambda v_i = (ay_{\bar{x}})_{x,\,i} - d_i v_i = -\rho_i,\quad 1 \leqslant i \leqslant N-1,$$

$$v_0 = v_N = 0.$$

OBSERVACIÓN 1. Si y_i se anula solamente en un extremo de la red $\bar{\omega}$, por ejemplo $y_N = 0$, entonces es cierta la estimación

$$\gamma_1(\rho y,\ y) \leqslant (ay_{\bar{x}}^2,\ 1)_{\omega+} + (dy,\ y) + \varkappa_0 y_0^2, \tag{31}$$

donde $1/\gamma_1 = \max\limits_{0\leqslant i\leqslant N-1} v_i$, y la función v_i es la solución del problema

$$\Lambda v_i = -\rho_i,\quad 0 \leqslant i \leqslant N-1,\quad v_N = 0,$$

$$\Lambda y_i = \begin{cases} \frac{2}{h}(a_1 y_{x,\,0} - \varkappa_0 y_0) - d_0 y_0, & i = 0, \\ (ay_{\bar{x}})_{x,\,i} - d_i y_i, & 1 \leqslant i \leqslant N-1,\quad \varkappa_0 \geqslant 0. \end{cases} \tag{32}$$

OBSERVACIÓN 2. Para la función reticular arbitraria y_i definida sobre $\bar{\omega}$, se puede obtener la estimación

$$\gamma_1(\rho y,\ y) \leqslant (ay_{\bar{x}}^2,\ 1)_{\omega+} + (dy,\ y) + \varkappa_0 y_0^2 + \varkappa_1 y_N^2, \tag{33}$$

donde $\varkappa_0 \geqslant 0$, $\varkappa_1 \geqslant 0$, $\varkappa_0 + \varkappa_1 + (d,\ 1) > 0$, y las funciones reticulares $\rho_i \geqslant 0$, $d_i \geqslant 0$ están definidas sobre $\bar{\omega}$. Aquí $1/\gamma_1 = \max\limits_{0\leqslant i\leqslant N} v_i$, donde v_i es la solución del problema de contorno

$$\Lambda v_i = -\rho_i,\quad 0 \leqslant i \leqslant N,$$

$$\Lambda y_i = \begin{cases} \frac{2}{h}(a_1 y_{x,\,0} - \varkappa_0 y_0) - d_0 y_0, & i = 0, \\ (ay_{\bar{x}})_{x,\,i} - d_i y_i, & 1 \leqslant i \leqslant N-1, \\ -\frac{2}{h}(a_N y_{\bar{x},\,N} + \varkappa_1 y_N) - d_N y_N, & i = N. \end{cases} \tag{34}$$

La demostración del lema 14 y de las observaciones 1 y 2 se realiza igual que la del lema 13. Aquí se aplica la función de Green de los operadores de diferencias Λ, indicados, la cual satisface las propiedades 1)—4) enumeradas más arriba.

LEMA 15. *Para la función reticular y_i, que se anula para $i = N$ es válida la estimación*

$$y_0^2 \leqslant th(\varepsilon l)\left[\varepsilon(y, y) + \frac{1}{\varepsilon}(y_{\bar{x}}^2, 1)_{\omega^+}\right], \quad \varepsilon \geqslant 0. \qquad (35)$$

La estimación análoga

$$y_N^2 \leqslant th(\varepsilon l)\left[\varepsilon(y, y) + \frac{1}{\varepsilon}(y_{\bar{x}}^2, 1)_{\omega^+}\right], \quad \varepsilon \geqslant 0,$$

es cierta para el caso, cuando $y_0 = 0$. Para la función reticular arbitraria y_i, definida sobre la red $\bar{\omega}$, tiene lugar la estimación

$$y_0^2 + y_N^2 \leqslant \frac{8+\varepsilon^2 l^2}{\varepsilon l\sqrt{16+\varepsilon^2 l^2}}\left[\varepsilon(y, y) + \frac{1}{\varepsilon}(y_{\bar{x}}^2, 1)_{\omega^+}\right], \quad \varepsilon > 0. \qquad (36)$$

Al principio demostremos la validez de la estimación (35). Para eso utilicemos la observación 1 al lema 14. Pongamos en (32) $a_i = 1/\varepsilon$, $d_i = \varepsilon$, $\varkappa_0 = 0$ y $\rho_0 = 2/h$, $\rho_i = 0$, $1 \leqslant i \leqslant N-1$. Entonces de (31) obtenemos la estimación

$$y_0^2 \leqslant \max_{0 \leqslant i \leqslant N-1} v_i\left[\varepsilon(y, y) + \frac{1}{\varepsilon}(y_{\bar{x}}^2, 1)_{\omega^+}\right],$$

donde v_i es la solución del siguiente problema auxiliar:

$$\begin{aligned} \Lambda v_i &= \frac{1}{\varepsilon} v_{\bar{x}x, i} - \varepsilon v_i = 0, \quad 1 \leqslant i \leqslant N-1, \\ \Lambda v_0 &= \frac{2}{\varepsilon h} v_{x, 0} - \varepsilon v_0 = -\frac{2}{h}, \quad v_N = 0. \end{aligned} \qquad (37)$$

Anotemos (37) por puntos

$$\begin{aligned} v_{i-1} &= 2\alpha v_i + v_{i+1} = 0, \quad 1 \leqslant i \leqslant N-1, \\ v_1 &- \alpha v_0 = -\varepsilon h, \quad v_N = 0, \end{aligned} \qquad (38)$$

donde $\alpha = 1 + 0{,}5\varepsilon^2 h^2 \geqslant 1$.

Nosotros obtuvimos el problema de contorno para una ecuación de diferencias de segundo orden con coeficientes constantes.

Utilizando la teoría general, desarrollada en el punto 1 del § 4 del cap. I, y también las propiedades de los polinomios de Chebishev (véase además el punto 2), hallaremos que la función

$$v_i = \frac{\varepsilon h U_{N-i-1}(\alpha)}{T_N(\alpha)}, \quad 0 \leqslant i \leqslant N$$

es la solución del problema (38). Aquí

$$T_n(\alpha) = \text{ch}(n \text{ Arch } \alpha),$$

$$U_n(\alpha) = \frac{\text{sh}((n+1)\text{ Arch } \alpha)}{\text{sh}(\text{Arch } \alpha)}, \quad |\alpha| \geqslant 1,$$

son los polinomios de Chebishev de grado n de primero y segundo género.

Puesto que $\alpha \geqslant 1$, entonces

$$\max_{0\leqslant i\leqslant N-1} v_i = v_0 = \frac{\varepsilon h U_{N-1}(\alpha)}{T_N(\alpha)}.$$

Así, está obtenida la estimación

$$y_0^2 \leqslant v_0 \left[\varepsilon (y, y) + \frac{1}{\varepsilon} (y_{\bar{x}}^2, 1)_{\omega^+} \right]$$

para la función reticular y_i, que satisface la condición $y_N = 0$. Esta estimación es exacta en el sentido, que ella pasa a ser igualdad, si en calidad de y_i tomamos la función v_i.

Estimemos ahora v_0 por encima para cualquier h. Si designamos $\operatorname{ch} 2z = \alpha$, entonces $z \geqslant 0$ y

$$\varepsilon h = 2 \operatorname{sh} z, \quad N = l/h = \varepsilon l/(2 \operatorname{sh} z),$$

$$T_N(\alpha) = \operatorname{ch} 2Nz = \operatorname{ch} w(z), \tag{39}$$

$$U_{N-1}(\alpha) = \frac{\operatorname{sh} 2Nz}{\operatorname{sh} 2z} = \frac{\operatorname{sh} w(z)}{2 \operatorname{sh} z \operatorname{ch} z}, \qquad w(z) = \frac{\varepsilon l_z}{\operatorname{sh} z}.$$

Por eso

$$v_0 = \frac{\operatorname{sh} w(z)}{\operatorname{ch} z \operatorname{ch} w(z)}.$$

Ya que para ε fijo

$$\frac{dw}{dz} = \frac{\varepsilon l (\operatorname{sh} z - z \operatorname{ch} z)}{\operatorname{sh}^2 z} \leqslant 0,$$

entonces

$$\frac{dv_0}{dz} = \frac{\operatorname{ch} z \dfrac{dw}{dz} - \operatorname{sh} z \operatorname{sh} w \operatorname{ch} w}{\operatorname{ch}^2 z \operatorname{ch}^2 w} \leqslant 0.$$

Por consiguiente, v_0 es máximo para $z = 0$. Esto da la estimación $v_0 \leqslant th(\varepsilon l)$. La desigualdad (35) está demostrada.

Ahora sea y_i la función reticular arbitraria. De la observación 2 al lema 14, cuando $a_i \equiv 1/\varepsilon$, $d_i \equiv \varepsilon$, $\varkappa_0 = \varkappa_1 = 0$, $\rho_0 = \rho_N = 2/h$, $\rho_i = 0$ para $1 \leqslant i \leqslant N-1$ obtenemos la estimación

$$y_0^2 + y_N^2 \leqslant \max_{0\leqslant i\leqslant N} v_i \left[\varepsilon (y, y) + \frac{1}{\varepsilon} (y_{\bar{x}}^2, 1)_{\omega^+} \right],$$

donde v_i es la solución del problema de contorno

$$\frac{1}{\varepsilon} v_{\bar{x}x, i} - \varepsilon v_i = 0, \quad 1 \leqslant i \leqslant N-1,$$

$$\frac{2}{\varepsilon h} v_{x,0} - \varepsilon v_0 = -\frac{2}{h}, \qquad -\frac{2}{\varepsilon h} v_{\bar{x},N} - \varepsilon v_N = -\frac{2}{h}.$$

La solución del problema (40) es la función

$$v_i = \frac{\varepsilon h [T_{N-i}(\alpha) + T_i(\alpha)]}{(\alpha^2 - 1) U_{N-1}(\alpha)}, \quad 0 \leqslant i \leqslant N,$$

donde α está definido más arriba.

De aquí encontramos que

$$\max_{0\leqslant i\leqslant N} v_i = v_0 = v_N = \frac{\varepsilon h\,(1+T_N(\alpha))}{(\alpha^2-1)\,N_{N-1}(\alpha)}. \qquad (41)$$

Estimemos esta expresión por encima para cualquier h. Utilizando (39) obtenemos

$$v_0 = \frac{1+\operatorname{ch} w(z)}{\operatorname{ch} z \operatorname{sh} w(z)} = \frac{\operatorname{ch}\frac{1}{2}w(z)}{\operatorname{ch} z \operatorname{sh}\frac{1}{2}w(z)} \leqslant \frac{\operatorname{ch}\frac{1}{2}w(z)}{\operatorname{sh}\frac{1}{2}w(z)} = \varphi(z).$$

Ya que

$$\frac{d\varphi}{dz} = -\frac{1}{\operatorname{sh}^2 0{,}5w}\,\frac{\partial w}{\partial z} > 0,$$

entonces la función $\varphi(z)$ es máxima para el $z = z_0$ máximo, el cual se encuentra de la relación $\operatorname{ch} 2z_0 = 1 + \varepsilon^2 l^2/8$ ($h \leqslant l/2$). De (39) obtenemos, que $w(z_0) = 4z_0$. Por lo tanto,

$$\varphi(z_0) = \frac{\operatorname{ch} 2z_0}{\operatorname{sh} 2z_0} = \frac{1+\varepsilon^2 l^2/8}{\sqrt{\varepsilon^2 l^2/8+\varepsilon^4 l^4/64}} = \frac{8+\varepsilon^2 l^2}{\varepsilon l\sqrt{16+\varepsilon^2 l^2}}.$$

La estimación (36) ha sido obtenida.

Los lemas 13 y 14 se generalizan sin dificultad al caso de una red no uniforme $\bar{\omega}$ arbitraria. En este caso para los productos escalares se utilizan las notaciones (4), (6), y los operadores de diferencias Λ se cambian por los correspondientes operadores sobre la red no uniforme.

LEMA 16. *Sean $\rho_i \geqslant 0$, $d_i \geqslant 0$ definidas sobre la red no uniforme arbitraria $\bar{\omega}$, $\rho_i \not\equiv 0$ y $a_i \geqslant c_1 > 0$ definida sobre ω^+. Sean $\varkappa_0 \geqslant 0$, $\varkappa_1 \geqslant 0$ números arbitrarios y supongamos que se cumple la condición $\varkappa_0 + \varkappa_1 + (d, \bar{l}) > 0$. Para cualquier función reticular y_i, definida sobre $\bar{\omega}$, es válida la desigualdad (33), donde $1/\gamma_1 = \max\limits_{0\leqslant i\leqslant N} v_i$, y v_i es la solución del problema $\Lambda v_i = -\rho_i$, $0 \leqslant i \leqslant N$. Aquí el operador Λ se define por las fórmulas*

$$\Lambda y_i = \begin{cases} \dfrac{1}{\hbar_0}(a_1 y_{x,0} - \varkappa_0 y_0) - d_0 y_0, & i = 0, \\ (a y_{\bar{x}})_{\hat{x},i} - d_i y_i, & 1 \leqslant i \leqslant N-1, \\ -\dfrac{1}{\hbar_N}(a_N y_{\bar{x},N} + \varkappa_1 y_N) - d_N y_N, & i = N. \end{cases} \qquad (42)$$

El lema 16 se demuestra igual que el lema anterior.

OBSERVACIÓN 1. Si $a_i \equiv 1$, $d_i \equiv 0$ y $\rho_i \equiv 1$, entonces la desigualdad (33) toma la forma

$$\gamma_1(y,\, y) \leqslant (y_{\bar{x}}^2,\, 1)_{\omega^+} + \varkappa_0 y_0^2 + \varkappa_1 y_N^2, \qquad (43)$$

donde

$$\gamma_1 = \frac{8(\varkappa_0 + \varkappa_1 + l\varkappa_0\varkappa_1)^2}{l(2+l\varkappa_0)(2+l\varkappa_1)(2\varkappa_0+2\varkappa_1+l\varkappa_0\varkappa_1)}.$$

Si, además, $y_0 = y_N = 0$, entonces la desigualdad (43) pasa a ser la desigualdad (30). Si y_i se anula solamente en un extremo, por ejemplo para $i = N$, entonces poniendo $y_N = 0$ en (43) y pasando al límite cuando $\varkappa_1 \to \infty$, obtenemos la estimación

$$\gamma_1 (y, y) \leqslant (y_{\bar{x}}^2, 1)_{\omega^+} + \varkappa_0 y_0^2, \quad \gamma_1 = \frac{8(1+l\varkappa_0)^2}{l^2(2+l\varkappa_0)^2}.$$

OBSERVACIÓN 2. De la definición (42) del operador en diferencias Λ y de la primera fórmula de Green de diferencias se deduce, que

$$(-\Lambda y, y) = (ay_{\bar{x}}^2, 1)_{\omega^+} + (dy, y) + \varkappa_0 y_0^2 + \varkappa_1 y_N^2.$$

Por eso la desigualdad (33) del lema (16) puede ser escrita en la forma

$$\gamma_1 (\rho y, y) \leqslant -(\Lambda y, y).$$

Pasemos a la deducción de la estimación (43). Hallemos la solución del problema $\Lambda v_i = -\rho_i$, $0 \leqslant i \leqslant N$, bajo las suposiciones indicadas en la observación 1. Tenemos el problema de contorno de diferencias

$$v_{\bar{x}x, i} = -1, \quad 1 \leqslant i \leqslant N-1, \tag{44}$$

$$v_{x, 0} = \varkappa_0 v_0 - \hbar_0, \quad i = 0, \tag{45}$$

$$-v_{\bar{x}, N} = \varkappa_1 v_N - \hbar_N, \quad i = N. \tag{46}$$

Multipliquemos la ecuación (44) por $\hbar_i$, sumemos según i desde j hasta $N-1$ y tengamos en cuenta la condición de contorno (46). Obtendremos

$$\sum_{i=j}^{N-1} v_{\bar{x}x, i} \hbar_i = \sum_{i=j}^{N-1} (v_{\bar{x}, i+1} - v_{\bar{x}, i}) = v_{\bar{x}, N} - v_{\bar{x}, j} =$$

$$= -\varkappa_1 v_N + \hbar_N - v_{\bar{x}, j} = -\sum_{i=j}^{N-1} \hbar_i = x_j - 0{,}5h_j - l + \hbar_N.$$

De aquí se deduce, que

$$v_{\bar{x}, j} = l - \varkappa_1 v_N + 0{,}5h_j - x_j, \quad 1 \leqslant j \leqslant N. \tag{47}$$

Poniendo $j = 1$ en (47) y teniendo en cuenta las igualdades $\hbar_0 = 0{,}5h_1$, $v_{\bar{x}, 1} = v_{x, 0} = \varkappa_0 v_0 - \hbar_0$, obtendremos la relación entre v_0 y v_N

$$\varkappa_0 v_0 + \varkappa_1 v_N = l. \tag{48}$$

Multiplicando (47) por h_j y sumando según j desde 1 hasta i, hallaremos

$$\sum_{j=1}^{i} v_{\bar{x},j} h_j = v_i - v_0 = (l - \varkappa_1 v_N) \sum_{j=1}^{i} h_j - \sum_{j=1}^{i} (x_j - 0,5h_j)\, h_j .$$

Como $h_j = x_j - x_{j-1}$, $x_j = 0,5h_j + 0,5\,(x_j + x_{j-1})$, entonces

$$\sum_{j=1}^{i} h_j = x_i, \quad \sum_{j=1}^{i} (x_j - 0,5h_j)\, h_j = 0,5 \sum_{j=1}^{i} (x_j^2 - x_{j-1}^2) = 0,5x_i^2 .$$

De esta forma tenemos

$$v_i = v_0 + x_i\,(l - \varkappa_1 v_N) - 0,5x_i^2 = v_0 +$$
$$+ 0,5\,(l - \varkappa_1 v_N)^2 - 0,5\,(x_i - l + \varkappa_1 v_N)^2, \quad 0 \leqslant i \leqslant N. \qquad (49)$$

Poniendo aquí $i = N$, hallaremos la segunda relación para v_0 y v_N

$$v_N = v_0 + l\,(l - \varkappa_1 v_N) - 0,5l^2. \qquad (50)$$

De (48) y (50) obtenemos

$$v_0 = \frac{l\,(2 + l\varkappa_1)}{2\,(\varkappa_0 + \varkappa_1 + \varkappa_0\varkappa_1 l)}, \quad v_N = \frac{l\,(2 + l\varkappa_0)}{2\,(\varkappa_0 + \varkappa_1 + \varkappa_0\varkappa_1 l)} . \qquad (51)$$

Puesto que $0 \leqslant l - \varkappa_1 v_N < l$, entonces de (49) y (51) hallaremos, que

$$\max_{0 \leqslant i \leqslant N} v_i \leqslant v_0 + 0,5\,(l - \varkappa_1 v_N)^2 =$$
$$= \frac{l\,(2 + l\varkappa_0)(2 + l\varkappa_1)\,(2\varkappa_0 + 2\varkappa_1 + l\varkappa_0\varkappa_1)}{8\,(\varkappa_0 + \varkappa_1 + l\varkappa_0\varkappa_1)^2} .$$

De aquí y del lema 16 se desprende la estimación (43). Si $y_0 = y_N = 0$, entonces, poniendo $a_i \equiv 1$, $d \equiv 0$ y $\rho_i \equiv 1$ en (33) y pasando al límite en (43) cuando $\varkappa_0 \to \infty$ y $\varkappa_1 \to \infty$, obtenemos la estimación (30) con $\gamma_1 = 8/l^2$.

5. Estimaciones superiores para operadores de diferencias. Obtengamos ahora estimaciones superiores para algunos operadores de diferencias.

LEMA 17. *Para la función reticular arbitraria y_i, prefijada sobre la red no uniforme $\bar{\omega}$, es válida la estimación*

$$(ay_{\bar{x}}^2,\ 1)_{\omega^+} \leqslant \gamma_2\,(y,\ y), \qquad (52)$$

donde

$$\gamma_2 = \max\left[\frac{4a_1}{h_1^2},\ \frac{4a_N}{h_N^2},\ \max_{1 \leqslant i \leqslant N-1} \frac{2}{\hbar_i}\left(\frac{a_i}{h_i} + \frac{a_{i+1}}{h_{i+1}}\right)\right].$$

Si la red es uniforme, entonces

$$\gamma_2 = \frac{4}{h^2} \max\left[a_1,\ a_N,\ \max_{1 \leqslant i \leqslant N-1}\left(\frac{a_i + a_{i+1}}{2}\right)\right].$$

Si $y_0 = y_N = 0$, entonces $\gamma_2 = \max\limits_{1 \leqslant i \leqslant N-1} \dfrac{2}{\hbar_i}\left(\dfrac{a_i}{h_i} + \dfrac{a_{i+1}}{h_{i+1}}\right)$.

En efecto, tenemos

$$(ay_{\bar{x}}^2,\ 1)_{\omega^+} = \sum_{i=1}^{N} \frac{a_i (y_i - y_{i-1})^2}{h_i} =$$

$$= \sum_{i=1}^{N} \frac{a_i}{h_i} y_i^2 + \sum_{i=0}^{N-1} \frac{a_{i+1}}{h_{i+1}} y_i^2 - 2 \sum_{i=1}^{N} \frac{a_i}{h_i} y_i y_{i-1}.$$

Utilizando la desigualdad $2y_i y_{i-1} \leqslant y_i^2 + y_{i-1}^2$, obtendremos para $a_i > 0$, que

$$(ay_{\bar{x}}^2, 1)_{\omega^+} \leqslant \sum_{i=1}^{N} \frac{2a_i}{h_i} y_i^2 + \sum_{i=0}^{N-1} \frac{2a_{i+1}}{h_{i+1}} y_i^2 =$$

$$= \frac{2a_1}{h_1 \hbar_0} y_0^2 \hbar_0 + \frac{2a_N}{h_N \hbar_N} y_N^2 \hbar_N + \sum_{i=1}^{N-1} \frac{2}{\hbar_i} \left(\frac{a_i}{h_i} + \frac{a_{i+1}}{h_{i+1}} \right) y_i^2 \hbar_i .$$

Como $\hbar_0 = 0{,}5h_1$, $\hbar_N = 0{,}5h_N$ y $(y,\ y) = \sum_{i=0}^{N} y_i^2 \hbar_i$, entonces de aquí se desprende la estimación (52) con el valor indicado para γ_2. El lema 17 está demostrado.

LEMA 18. *Sean $a_i > 0$, $b_i \geqslant 0$, y σ_0, σ_1 no negativos, al mismo tiempo $(b,\ 1) + \sigma_0 + \sigma_1 \neq 0$. Para una función reticular arbitraria y_i, definida sobre la red no uniforme $\bar{\omega}$, es válida la estimación*

$$(ay_{\bar{x}}^2,\ 1)_{\omega^+} + (by,\ y) + \sigma_0 y_0^2 + \sigma_1 y_N^2 \leqslant \bar{\gamma}_2 (y,\ y), \qquad (53)$$

donde $\bar{\gamma}_2 = \gamma_2 + (1 + \gamma_2) \max\limits_{0 \leqslant i \leqslant N} v_i$, γ_2 está definido en el lema (17), y v es la solución del problema de contorno

$$(av_{\bar{x}})_{\hat{x},\, i} - v_i = -b_i, \quad 1 \leqslant i \leqslant N-1,$$

$$\frac{a_1}{\hbar_0} v_{x,\, 0} - v_0 = -b_0 - \frac{\sigma_0}{\hbar_0}, \quad i = 0, \qquad (54)$$

$$-\frac{a_N}{\hbar_N} v_{\bar{x},\, N} - v_N = -b_N - \frac{\sigma_1}{\hbar_N}, \quad i = N.$$

En efecto, del lema 16 para $\rho_i = b_i$ si $1 \leqslant i \leqslant N - 1$, $\rho_0 = b_0 + \sigma_0/\hbar_0$, $\rho_N = b_N + \sigma_1/\hbar_N$ y $\varkappa_0 = \varkappa_1 = 0$, $d_i \equiv 1$ obtenemos la estimación

$$(by,\ y) + \sigma_0 y_0^2 + \sigma_1 y_N^2 = (\rho y,\ y) \leqslant \max_{0 \leqslant i \leqslant N} v_i \left[(ay_{\bar{x}}^2,\ 1)_{\omega^+} + (y,\ y) \right],$$

donde v_i es la solución del problema auxiliar (54). Empleando el lema 17 tendremos

$$(ay_{\bar{x}}^2, 1)_{\omega^+} + (by, y) + \sigma_0 y_0^2 + \sigma_1 y_N^2 \leqslant (1+c)(ay_{\bar{x}}^2, 1)_{\omega^+} + c(y, y) \leqslant$$
$$\leqslant [\gamma_2 + (1+\gamma_2)c](y, y), \quad c = \max_{0 \leqslant i \leqslant N} v_i.$$

El lema 18 está demostrado.

6. **Esquemas de diferencias como ecuaciones operacionales en espacios abstractos.** Después del cambio de las derivadas, que entran en la ecuación diferencial y las condiciones de contorno, por las relaciones de diferencias sobre cierta red $\bar{\omega}$ nosotros obtenemos un esquema de diferencias. Las ecuaciones en diferencias, que relacionan los valores buscados de la función reticular en los nodos de $\bar{\omega}$, forman un sistema de ecuaciones algebraicas. Este sistema es lineal, si el problema inicial es lineal.

El esquema de diferencias se define por el operador de diferencias, el cual define la estructura de las ecuaciones en diferencias en los nodos de la red, donde se busca la solución incógnita, y por las condiciones de contorno en los nodos de la frontera. El operador de diferencias actúa en el espacio de las funciones reticulares, prefijadas sobre $\bar{\omega}$.

Examinemos un ejemplo. Supongamos que es necesario hallar la solución del problema

$$u'' = -\varphi(x), \quad 0 < x < l,$$
$$u'(0) = \varkappa_0 u(0) - \mu_1, \quad u(l) - \mu_2, \quad \varkappa_0 \geqslant 0. \tag{55}$$

sobre el intervalo $0 \leqslant x \leqslant l$.

Sobre la red uniforme $\bar{\omega} = \{x_i - ih, \ i = 0, 1, \ldots, N, \ hN = l\}$ ponemos en correspondencia al problema (55) el esquema de diferencias

$$\Lambda y_i = y_{\bar{x}x, i} = -\varphi_i, \quad 1 \leqslant i \leqslant N-1,$$
$$\Lambda y_0 = \frac{2}{h}(y_{x,0} - \varkappa_0 y_0) = -\left(\varphi_0 + \frac{2}{h}\mu_1\right), \tag{56}$$
$$y_N = \mu_2.$$

El operador de diferencias Λ está definido sobre un conjunto $(N+1)$-dimensional de funciones reticulares, definidas sobre $\bar{\omega}$, y lo aplica sobre un conjunto N-dimensional de funciones, definidas en $\omega^- = \{x_i \in \bar{\omega}, \ i = 0, 1, \ldots, N-1\}$. Está claro, que el dominio de definición y el campo de valores del operador Λ no coinciden.

Examinemos ahora el espacio $H(\omega^-)$ de las funciones reticulares, prefijadas sobre ω^-. Definiremos el producto escalar en $H(\omega^-)$, como en el ejemplo 1 del punto 1 del § 2:

$$(u, v) = \sum_{i=1}^{N-1} u_i v_i h + 0{,}5 h u_0 v_0, \quad u, v \in H(\omega^-).$$

Definamos ahora el operador lineal A de la siguiente forma: $Ay_i = -\Lambda \overset{\circ}{y}_i$, $0 \leqslant i \leqslant N-1$, donde $y \in H(\omega^-)$, $\overset{\circ}{y}_i = y_i$ para $0 \leqslant i \leqslant N-1$ y $\overset{\circ}{y}_N = 0$. Aprovechando esta definición, daremos la escritura detallada del operador A:

$$Ay_i = \begin{cases} -\frac{2}{h}(y_{x,0} - \varkappa_0 y_0), & i = 0, \\ -y_{\bar{x}x,i}, & 1 \leqslant i \leqslant N-2, \\ \frac{1}{h^2}(2y_{N-1} - y_{N-2}), & i = N-1. \end{cases} \tag{57}$$

El operador A aplica $H(\omega^-)$ sobre $H(\omega^-)$ y es lineal.

Transformemos el esquema de diferencias (2). Teniendo en cuenta la condición $y_N = \mu_2$, escribamos (56) en la forma

$$\begin{aligned} &-\frac{2}{h}(y_{x,0} - \varkappa_0 y_0) = f_0 = \left(\varphi_0' + \frac{2}{h}\mu_1\right), \\ &-y_{\bar{x}x,i} = f_i = \varphi_i, \quad 1 \leqslant i \leqslant N-2, \\ &\frac{1}{h^2}(2y_{N-1} - y_{N-2}) = f_{N-1} = \left(\varphi_{N-1} + \frac{1}{h^2}\mu_2\right). \end{aligned} \tag{58}$$

Comparando (57) y (58), hallaremos, que el esquema de diferencias (56) se escribe en forma de la ecuación operacional de primer género

$$Ay = f, \tag{59}$$

donde y es la incógnita, f es el elemento dado del espacio $H(\omega^-)$, y A es el operador, que actúa en $H(\omega^-)$, definido más arriba.

Indiquemos las propiedades fundamentales del operador A.

El operador A es autoconjugado en $H(\omega^-)$, es decir

$$(Au, v) = (u, Av), \quad u, v \in H(\omega^-).$$

En efecto, $(Au, v) = -(\Lambda\overset{\circ}{u}, \overset{\circ}{v})$ y al mismo tiempo $\overset{\circ}{u}_N = \overset{\circ}{v}_N = 0$. Aplicando la segunda fórmula de Green de

diferencias (13), obtenemos

$$(\Lambda \overset{\circ}{u}, \overset{\circ}{v}) = \sum_{i=1}^{N-1} \overset{\circ}{u}_{\bar{x}x,\,i} \overset{\circ}{v}_i h + (\overset{\circ}{u}_{x,\,0} - \varkappa_0 \overset{\circ}{u}_0)\, \overset{\circ}{v}_0 =$$

$$= \sum_{i=1}^{N-1} \overset{\circ}{u}_i \overset{\circ}{v}_{\bar{x}x,\,i} h + (\overset{\circ}{u}_{\bar{x}} \overset{\circ}{v} - \overset{\circ}{v}_{\bar{x}} \overset{\circ}{u})_N - (\overset{\circ}{u}_x \overset{\circ}{v} - \overset{\circ}{v}_x \overset{\circ}{u})_0 +$$

$$+ (\overset{\circ}{u}_x \overset{\circ}{v} - \varkappa_0 \overset{\circ}{u} \overset{\circ}{v})_0 = \sum_{i=1}^{N-1} \overset{\circ}{u}_i \overset{\circ}{v}_{\bar{x}x,\,i} h + (\overset{\circ}{v}_x \overset{\circ}{u} - \varkappa_0 \overset{\circ}{v} \overset{\circ}{u})_0 = (\overset{\circ}{u}, \Lambda \overset{\circ}{v}).$$

La afirmación está demostrada.

El operador A está definido positivamente, es decir

$$(Au, u) \geqslant \gamma_1 (u, u), \quad u \in H(\omega^-),$$

donde $\gamma_1 = \dfrac{8(1+l\varkappa_0)^2}{l^2(2+l\varkappa_0)^2} \geqslant \dfrac{2}{l^2} > 0$. Esta afirmación se deduce de las observaciones 1 y 2 al lema 16. El operador A en virtud del lema 10 tiene un operador inverso acotado A^{-1}. Por eso la solución de la ecuación (59) existe y es única.

Para el operador A tiene lugar la estimación superior

$$(Au, u) \leqslant \gamma_2 (u, u), \quad u \in H(\omega^-),$$

donde $\gamma_2 = \dfrac{4}{h^2}\left(1 + \varkappa_0 \dfrac{h}{2}\right)$, ya que $y_N = 0$ y

$$(Ay, y) = (y_{\bar{x}}^2, 1)_{\omega^+} + \varkappa_0 y_0^2,$$

$$y_0^2 \leqslant \frac{2}{h}(y, y), \quad (y_{\bar{x}}^2, 1)_{\omega^+} \leqslant \frac{4}{h^2}.$$

La última desigualdad se deduce del lema 17.

En calidad de segundo ejemplo examinemos sobre la red no uniforme

$$\bar{\omega} = \{x_i \in [0, l], \ x_i = x_{i-1} + h_i, \ 1 \leqslant i \leqslant N, \ x_0 = 0, \ x_N = l\}$$

el esquema de diferencias

$$\Lambda y_i = (a y_{\bar{x}})_{\hat{x},\,i} - d_i y_i = -\varphi_i, \quad 1 \leqslant i \leqslant N-1,$$

$$\Lambda y_0 = \frac{1}{\hbar_0}(a_1 y_{x,\,0} - \varkappa_0 y_0) - d_0 y_0 = -\left(\varphi_0 + \frac{1}{\hbar_0}\mu_1\right), \quad i = 0, \tag{60}$$

$$\Lambda y_N = -\frac{1}{\hbar_N}(a_N y_{\bar{x},\,N} + \varkappa_1 y_N) - d_N y_N =$$

$$= -\left(\varphi_N + \frac{1}{\hbar_N}\mu_2\right), \quad i = N.$$

El esquema (60) aproxima el tercer problema de contorno para la ecuación con coeficientes variables

$$(ku')' - qu = -\varphi(x), \quad 0 < x < l,$$
$$ku' = \varkappa_0 u - \mu_1, \quad x = 0,$$
$$-ku' = \varkappa_1 u - \mu_2, \quad x = l$$

para la correspondiente elección de los coeficientes a_i y d_i, por ejemplo para $a_i = k(x_i - 0{,}5h_i)$ y $d_i = q(x_i)$.

Si en el espacio $H(\bar{\omega})$ de las funciones escalares, definidas sobre $\bar{\omega}$, con el producto escalar

$$(u, v) = \sum_{i=0}^{N} u_i v_i \hbar_i, \quad \hbar_0 = 0{,}5h_1, \quad \hbar_N = 0{,}5h_N,$$

definimos el operador $A = -\Lambda$ y la función reticular $f_i = \varphi_i$, $1 \leqslant i \leqslant N-1$, $f_0 = \varphi_0 + \mu_1/\hbar_0$, $f_N = \varphi_N + \mu_2/\hbar_N$, entonces el esquema de diferencias (60) se escribirá en forma de ecuación operacional (59).

La autoconjugación del operador A, que aplica $H(\bar{\omega})$ sobre $H(\bar{\omega})$, se desprende de la segunda fórmula de Green de diferencias. Si se cumplen las condiciones $a_i \geqslant c_1 > 0$, $d_i \geqslant 0$, $\varkappa_0 \geqslant 0$, $\varkappa_1 \geqslant 0$ y $\varkappa_0 + \varkappa_1 + (d, 1) > 0$, entonces el operador A, es definido positivo en $H(\bar{\omega})$, y es válida la estimación $(Au, u) \geqslant \gamma_1 (u, u)$, $1/\gamma_1 = \max\limits_{0 \leqslant i \leqslant N} v_i$, donde v_i es la solución del problema $\Lambda v_i = -1$, $0 \leqslant i \leqslant N$. Notemos, que la positividad de v_i se deduce del principio del máximo, cierto para el operador Λ bajo las condiciones indicadas.

Si $d_i \equiv 0$, entonces se puede obtener un estimado burdo para γ_1 de la siguiente forma. De la primera fórmula de Green de diferencias obtenemos

$$(Ay, y) = (-\Lambda y, y) = (ay_{\bar{x}}^2, 1)_{\omega^+} + \varkappa_0 y_0^2 + \varkappa_1 y_1^2.$$

En virtud de las condiciones $a_i \geqslant c_i > 0$, $1 \leqslant i \leqslant N$, de aquí hallaremos

$$(Ay, y) \geqslant c_1 [(y_{\bar{x}}^2, 1)_{\omega^+} + \bar{\varkappa}_0 y_0^2 + \bar{\varkappa}_1 y_1^2],$$

donde $c_1 \bar{\varkappa}_0 = \varkappa_0$, $c_1 \bar{\varkappa}_1 = \varkappa_1$. Ya que $\varkappa_0 + \varkappa_1 > 0$, entonces de la observación 1 al lema 16 obtenemos la estimación

$$(y_{\bar{x}}^2, 1)_{\omega^+} + \bar{\varkappa}_0 y_0^2 + \varkappa_1 y_1^2 \geqslant \bar{\gamma}_1 (y, y),$$

donde

$$\overline{\gamma}_1 = \frac{8(\overline{\varkappa}_0 + \overline{\varkappa}_1 + l\overline{\varkappa}_0\overline{\varkappa}_1)^2}{l(2 + l\overline{\varkappa}_0)(2 + l\overline{\varkappa}_1)(2\overline{\varkappa}_0 + 2\overline{\varkappa}_1 + l\overline{\varkappa}_0\overline{\varkappa}_1)}.$$

Sustituyendo aquí $\overline{\varkappa}_0$ y $\overline{\varkappa}_1$, hallaremos que $(Au,\ u) \geqslant \gamma_1 (u,\ u)$, donde

$$\gamma_1 = c_1\overline{\gamma}_1 = \frac{8c_1\,(c_1\varkappa_0 + c_1\varkappa_1 + l\varkappa_0\varkappa_1)^2}{l\,(2c_1 + l\varkappa_0)\,(2c_1 + l\varkappa_1)\,(2c_1\varkappa_0 + 2c_1\varkappa_1 + l\varkappa_0\varkappa_1)}.$$

Para el operador A tiene lugar la estimación por encima $(Au,\ u) \leqslant \gamma_2\ (u,\ u)$, donde γ_2 está definido en el lema 18, ya que

$$(Ay,\ y) = (ay_{\bar{x}}^2,\ 1)_{\omega^+} + (dy^2,\ 1) + \varkappa_0 y_0^2 + \varkappa_1 y_N^2.$$

En el ejemplo examinado el operador A y el operador de diferencias Λ están definidos en el mismo espacio de funciones reticulares $H(\overline{\omega})$ y se diferencian sólo por el signo. A diferencia del primer ejemplo, los miembros derechos del esquema de diferencias (60) y de la ecuación operacional (59) coinciden.

Nosotros nos limitamos aquí a los ejemplos más simples. En el próximo punto por un procedimiento análogo se reducirán a las ecuaciones operacionales los esquemas de diferencias, que aproximan los problemas de contorno elípticos en un espacio de varias dimensiones, en los correspondientes espacios de Hilbert de dimensión finita de funciones reticulares. Serán también estudiadas las propiedades fundamentales de tales operadores.

De los ejemplos citados se desprende, que los esquemas de diferencias se pueden interpretar como ecuaciones operacionales con operadores en un espacio lineal normado de dimensión finita. Para estos operadores es característico, que ellos apliquen todo el espacio en sí mismo.

7. Esquemas de diferencias para ecuaciones elípticas de coeficientes constantes. Sean $\overline{G} = \{0 \leqslant x_\alpha \leqslant l_\alpha,\ \alpha = 1,\ 2\}$ un rectángulo, $\overline{\omega} = \{x_{ij} = (ih_1,\ jh_2) \in \overline{G},\ 0 \leqslant i \leqslant N_1,\ 0 \leqslant \leqslant j \leqslant N_2,\ h_\alpha N_\alpha = l_\alpha,\ \alpha = 1,\ 2\}$ una red en $\overline{G}$, y γ el conjunto de los nodos frontera de la red $\overline{\omega}$. La red es uniforme por cada dirección x_α y su paso es h_α. Designemos mediante ω el conjunto de los nodos interiores de la red. Introduzcamos el espacio de las funciones reticulares $H = H(\omega)$, definidas sobre ω. Definiremos en H el producto

escalar

$$(u, v)=\sum_{i=1}^{N_1-1}\sum_{j=1}^{N_2-1} u(i, j)\, v(i, j)\, h_1 h_2.$$

Examinemos el problema de Dirichlet de diferencias para la ecuación de Poisson en la red $\bar{\omega}$

$$\Lambda y=\sum_{\alpha=1}^{2} \Lambda_\alpha y=-\varphi(x), \quad x \in \omega,$$

$$y(x)=g(x), \quad x \in \gamma, \tag{61}$$

donde $\Lambda_\alpha y=y_{\bar{x}_\alpha x_\alpha}$, $\alpha=1, 2$.

El esquema de diferencias (61) se puede escribir en la forma de la ecuación operacional (59). Para esto definamos el operador A mediante la fórmula $Ay=-\Lambda \overset{\circ}{y}$, $x \in \omega$, donde $y \in H$, $\overset{\circ}{y} \in \overset{\circ}{H}$ y $y(x)=\overset{\circ}{y}(x)$ para $x \in \omega$. Aquí $\overset{\circ}{H}$ es el conjunto de funciones reticulares, definidas sobre $\bar{\omega}$ y que se anulan sobre γ. El segundo miembro f de la ecuación (59) se diferencia del segundo miembro φ del esquema de diferencias (61) solamente en los nodos fronterizos

$$f=\varphi+\varphi_1/h_1^2+\varphi_2/h_2^2,$$

donde

$$\varphi_1(x)=\begin{cases} g(0, x_2), & x_1=h_1,\\ 0, & 2h_1\leqslant x_1\leqslant l_1-2h_1,\\ g(l_1, x_2), & x_1=l_1-h_1,\end{cases}$$

$$\varphi_2(x)=\begin{cases} g(x_1, 0), & x_2=h_2,\\ 0, & 2h_2\leqslant x_2\leqslant l_2-2h_2,\\ g(x_1, l_2), & x_2=l_2-h_2.\end{cases}$$

Investiguemos las propiedades del operador A, que actúa de $H(\omega)$ en $H(\omega)$.

1. El operador A es autoconjugado:

$$(Au, v)=(u, Av), \quad u, v \in H(\omega). \tag{62}$$

Para la demostración tengamos en cuenta que

$$(A_1u, v)=(-\Lambda_1\overset{\circ}{u}, \overset{\circ}{v})=-\sum_{j=1}^{N_2-1} h_2\sum_{i=1}^{N_1-1} h_1(\overset{\circ}{v}\Lambda_1\overset{\circ}{u})_{ij}=$$

$$=-\sum_{j=1}^{N_2-1} h_2\sum_{i=1}^{N_1-1} h_1(\overset{\circ}{u}\Lambda_1\overset{\circ}{v})_{ij}=-(\overset{\circ}{u}, \Lambda_1\overset{\circ}{v})=(u, A_1v),$$

ya que en virtud de la segunda fórmula de Green de diferencias, sobre la red $\bar{\omega}_1 = \{x_i(i) = ih_1, 0 \leqslant i \leqslant N_1, h_1N_1 = l_1\}$, el operador de diferencias Λ_1 satisface la igualdad

$$\sum_{i=1}^{N_1-1} h_1(\overset{\circ}{v}\Lambda_1\overset{\circ}{u})_{ij} = \sum_{i=1}^{N_1-1} h_1(\overset{\circ}{u}_1\Lambda_1\overset{\circ}{v})_{ij}$$

y además, se puede cambiar el orden de sumación por i y j.

Análogamente encontramos, que $(A_2u, v) = (u, A_2v)$. De aquí se deduce (62).

2. El operador A es definido positivo, y para él son válidos las estimaciones

$$\delta E \leqslant A \leqslant \Delta E, \quad \delta > 0, \tag{63}$$

donde

$$\delta = \sum_{\alpha=1}^{2} \frac{2}{h_\alpha^2} \operatorname{sen}^2 \frac{\pi}{2N_\alpha} \geqslant \sum_{\alpha=1}^{2} \frac{8}{l_\alpha^2},$$

$$\Delta = \sum_{\alpha=1}^{2} \frac{4}{h_\alpha^2} \cos^2 \frac{\pi}{2N_\alpha} \leqslant \sum_{\alpha=1}^{2} \frac{4}{h_\alpha^2}. \tag{64}$$

Notemos, que δ y Δ son los valores propios mínimo y máximo del operador de Laplace Λ de diferencias (véase el punto 1, § 2, cap. IV).

Esta afirmación se demuestra igual que el lema 12. De esta forma hemos establecido, que en $H = H(\omega)$

$$A = A^*, \quad \delta E \leqslant A \leqslant \Delta E, \quad \delta > 0.$$

Si sobre una parte γ_0 de la frontera reticular γ se prefija la condición de contorno de primer género $y(x) = g(x)$, $x \in \gamma_0$, y sobre la parte restante se dan condiciones de contorno de segundo o tercer género, entonces el operador A se define por el método descrito más arriba, al mismo tiempo $\overset{\circ}{H}$ es el conjunto de las funciones que se anulan solamente sobre γ_0, y $H = H(\omega_0)$ es el espacio de las funciones reticulares, definidas sobre $\omega_0 = \omega \cup (\gamma \setminus \gamma_0)$. Por ejemplo, sea $\gamma_0 = \{x_{ij} \in \omega, i = 0, 0 \leqslant j \leqslant N_2\}$, y supongamos que sobre $\gamma \setminus \gamma_0$ están dadas condiciones de contorno de segundo género. Entonces el esquema de diferencias se escribe en la forma

$$\Lambda y = (\Lambda_1 + \Lambda_2) y = -\varphi(x), \quad x \in \omega_0,$$
$$y(x) = g(x), \quad x \in \gamma_0.$$

Aquí

$$\Lambda_2 y = \begin{cases} \frac{2}{h_2} y_{x_2}, & x_2 = 0, \\ y_{\bar{x}_2 x_2}, & h_2 \leqslant x_2 \leqslant l_2 - h_2, \\ -\frac{2}{h_2} y_{\bar{x}_2}, & x_2 = l_2, \quad h_1 \leqslant x_1 \leqslant l_1, \end{cases}$$

y el operador Λ_1 se da por las fórmulas

$$\Lambda_1 y = \begin{cases} y_{\bar{x}_1 x_1}, & h_1 \leqslant x_1 \leqslant l_1 - h_1, \\ -\frac{2}{h_1} y_{\bar{x}_1}, & x_1 = l_1, \quad 0 \leqslant x_2 \leqslant l_2. \end{cases}$$

El producto escalar en el espacio $H = H(\omega_0)$ es define por la fórmula

$$(u, v) = \sum_{i=1}^{N_1} \sum_{j=0}^{N_2} u(i, j) v(i, j) \hbar_1(i) \hbar_2(j),$$

donde

$$\hbar_1(i) = \begin{cases} h_1, & 1 \leqslant i \leqslant N_1 - 1, \\ 0{,}5h_1, & i = N_1, \end{cases}$$

$$\hbar_2(j) = \begin{cases} h_2, & 1 \leqslant j \leqslant N_2 - 1, \\ 0{,}5h_2, & j = 0, \ N_2. \end{cases}$$

Se puede mostrar que el operador $A = A_1 + A_2$ correspondiente al operador de diferencias Λ es autoconjugado en H y para él son ciertas las estimaciones (63) con $\delta = \delta_1 + \delta_2$, $\Delta = \Delta_1 + \Delta_2$, $\delta_1 = \frac{4}{h_1^2} \operatorname{sen}^2 \frac{\pi}{4N_1}$, $\Delta_1 = \frac{4}{h_1^2} \cos^2 \frac{\pi}{4N_1}$, $\delta_2 = 0$ y $\Delta_2 = \frac{4}{h_2^2}$. Aquí δ_α y Δ_α son los valores propios mínimo y máximo del operador de diferencias Λ_α, $\alpha = 1, 2$.

Notemos, que los operadores A_1 y A_2 conmutan tanto para el primero como para el segundo problema de contorno. Por eso, en virtud de la teoría general (véase el punto 5, § 1, cap. V) los valores propios del operador A son la suma de los valores propios de los operadores A_1 y A_2: $\lambda(A) = \lambda(A_1) + \lambda(A_2)$.

8. **Ecuaciones de coeficientes variables y con derivadas mixtas.** Examinemos el problema de Dirichlet para una ecuación elíptica de coeficientes variables en el rectángulo

$\overline{G} = \{0 \leqslant x_\alpha \leqslant l_\alpha,\ \alpha = 1,\ 2\};$

$$Lu = \sum_{\alpha=1}^{2} \frac{\partial}{\partial x_\alpha}\left(k_\alpha (x) \frac{\partial u}{\partial x_\alpha} \right) - q(x)\, u = -\varphi (x), \quad x \in G, \qquad (65)$$

$u(x) = g(x), \quad x \in \Gamma,$

donde $k_\alpha (x)$ y $q(x)$ son las funciones suficientemente suaves que satisfacen las condiciones $0 < c_1 \leqslant k_\alpha(x) \leqslant c_2$, $0 \leqslant \leqslant d_1 \leqslant q(x) \leqslant d_2$. Designemos mediante $\overline{\omega} = \omega + \gamma$ la red con pasos h_1 y h_2 introducida en el punto 7.

Pongámosle en correspondencia al problema (65) el problema de Dirichlet de diferencias sobre la red $\overline{\omega}$:

$$\Lambda y = (\Lambda_1 + \Lambda_2)\, y - dy = -\varphi (x), \quad x \in \omega, \qquad (66)$$

$$y(x) = g(x), \quad x \in \gamma,$$

donde $\Lambda_\alpha y = (a_\alpha y_{\bar{x}_\alpha})_{x_\alpha}$, $\alpha = 1,\ 2$, y $a_\alpha (x)$, $d(x)$ se expresan por ejemplo, así:

$$a_1 (x_1,\ x_2) = k_1 (x_1 - 0{,}5h_1,\ x_2),$$

$$a_2 (x_1,\ x_2) = k_2 (x_1,\ x_2 - 0{,}5h_2), \quad d(x) = q(x).$$

Entonces los coeficientes del esquema de diferencias satisfacen las condiciones

$$0 < c_1 \leqslant a_\alpha (x) \leqslant c_2, \quad 0 \leqslant d_1 \leqslant d \leqslant d_2. \qquad (67)$$

Designemos mediante $H = H(\omega)$ el espacio de las funciones reticulares, introducido en el punto anterior, y mediante $\overset{\circ}{H}$ el conjunto de funciones reticulares, que se anulan sobre γ.

Escribamos el esquema de diferencias (66) en la forma de la ecuación operacional (59), donde definimos el operador A de la forma usual: $Ay = -\Lambda \overset{\circ}{y}$, siendo $y \in H$, $\overset{\circ}{y} \in \overset{\circ}{H}$ e $y(x) = \overset{\circ}{y}(x)$ para $x \in \omega$.

Designemos mediante $\mathcal{R} = \mathcal{R}_1 + \mathcal{R}_2$, donde $\mathcal{R}_\alpha y = -y_{\bar{x}_\alpha x_\alpha}$, $\alpha = 1,\ 2$ el operador de Laplace de diferencias y definamos en H su correspondiente operador R: $Ry = -\mathcal{R}\overset{\circ}{y}$, $y \in H$, $\overset{\circ}{y} \in \overset{\circ}{H}$ y $y(x) = \overset{\circ}{y}(x)$ para $x \in \omega$.

LEMA 10. *El operador A es autoconjugado en H, y para él son ciertas las estimaciones*

$$(c_1 + d_1/\Delta)\,(Ru,\ u) \leqslant (Au,\ u) \leqslant$$
$$\leqslant (c_2 + d_2/\delta)\,(Ru,\ u), \qquad (68)$$
$$(c_1\delta + d_1)\,(u,\ u) \leqslant (Au,\ u) \leqslant$$
$$\leqslant (c_2\Delta + d_2)\,(u,\ u), \qquad (69)$$

donde δ y Δ están definidos en (64).

En efecto, de las condiciones (67) y de las estimaciones obtenidas en el punto anterior

$$\delta E \leqslant R \leqslant \Delta E, \qquad (70)$$

se deduce, que para cualquier $u \in H$ son ciertas las desigualdades

$$\frac{d_1}{\Delta}(Ru,\ u) \leqslant d_1(u,\ u) \leqslant (du,\ u) \leqslant d_2(u,\ u) \leqslant \frac{d_2}{\delta}(Ru,\ u). \qquad (71)$$

Seguidamente, la primera fórmula de Green da

$$(A_1u,\ u) = -(\Lambda_1\overset{\circ}{u},\ \overset{\circ}{u}) = \sum_{j=1}^{N_2-1}\sum_{i=1}^{N_1}(a_1\overset{\circ}{u}{}^2_{\bar{x}_1})_{ij}\,h_1h_2,$$
$$(R_1u,\ u) = -(\mathcal{R}_1\overset{\circ}{u},\ \overset{\circ}{u}) = \sum_{j=1}^{N_2-1}\sum_{i=1}^{N_1}(\overset{\circ}{u}{}^2_{\bar{x}_1})_{ij}\,h_1h_2.$$

En virtud de (67) de aquí obtenemos las desigualdades

$$c_1\,(R_1u,\ u) \leqslant (A_1u,\ u) \leqslant C_2\,(R_1u,\ u).$$

Análogamente hallamos, que

$$c_1\,(R_2u,\ u) \leqslant (A_2u,\ u) \leqslant c_2\,(R_2u,\ u).$$

De aquí y de (70) se desprenden las desigualdades

$$c_1\delta\,(u,\ u) \leqslant c_1\,(Ru,\ u) \leqslant ((A_1 + A_2)\,u,\ u) \leqslant$$
$$\leqslant c_2\,(Ru,\ u) \leqslant c_2\Delta\,(u,\ u),$$

y sumándolas con las desigualdades (71) obtendremos (68) y (69).

La autoconjugación del operador A se demuestra análogamente al punto anterior.

Señalemos, que en las desigualdades (68) están indicadas las constantes de equivalencia energética de los operadores R y A, y al mismo tiempo, como $d_1 \geqslant 0$ y $\delta \geqslant 8/l_1^2 + 8/l_2^2$,

entonces estos operadores son equivalentes con constantes que no dependen del número de nodos de la red.

Examinemos ahora el *problema de Dirichlet para una ecuación elíptica que contiene derivadas mixtas*

$$Lu = \sum_{\alpha, \beta=1}^{2} \frac{\partial}{\partial x_\alpha}\left(k_{\alpha\beta}(x)\frac{\partial u}{\partial x_\beta}\right) = -\varphi(x), \quad x \in \overline{G}, \tag{72}$$

$$u(x) = g(x), \quad x \in \Gamma.$$

Se supone, que se cumplen las condiciones de elipticidad

$$c_1 \sum_{\alpha=1}^{2} \xi_\alpha^2 \leqslant \sum_{\alpha, \beta=1}^{2} k_{\alpha\beta}(x)\,\xi_\alpha\xi_\beta \leqslant c_2 \sum_{\alpha=1}^{2} \xi_\alpha^2, \quad x \in \overline{G}, \tag{73}$$

donde $c_2 \geqslant c_1 > 0$, y $\xi = (\xi_1, \xi_2)$ es un vector arbitrario.

Sobre la red rectangular $\overline{\omega}$ al problema (72) se le puede poner en correspondencia el esquema de diferencias

$$\Lambda y = 0{,}5 \sum_{\alpha, \beta=1}^{2} [(k_{\alpha\beta} y_{\bar{x}_\beta})_{x_\alpha} + (k_{\alpha\beta} y_{x_\beta})_{\bar{x}_\alpha}] = -\varphi(x), \; x \in \omega,$$

$$y(x) = g(x), \quad x \in \gamma. \tag{74}$$

Escribamos (74) en la forma de la ecuación operacional (59), definiendo el operador A de la manera usual: $Ay = -\Lambda \overset{\circ}{y}$, donde $y \in H(\omega)$, $\overset{\circ}{y} \in \overset{\circ}{H}$ e $\overset{\circ}{y}(x) = y(x)$ para $x \in \omega$. Con esto el miembro derecho f se diferencia del miembro derecho φ de la ecuación (74) solamente en los nodos fronterizos. Para encontrar la forma explícita de f se debe escribir la ecuación en diferencias en un nodo fronterizo, utilizar las condiciones de contorno y pasar al miembro derecho de la ecuación los valores conocidos de $y(x)$ sobre γ.

Mostremos ahora, que al cumplirse las condiciones de simetría $k_{12}(x) = k_{21}(x)$ el operador A es autoconjugado en el espacio $H = H(\omega)$, definido más arriba. Para esto escribamos el operador Λ en forma de la suma $\Lambda = (\Lambda_1 + \Lambda_2)/2$, donde

$$\Lambda_\alpha y = (k_{\alpha\alpha} y_{\bar{x}_\alpha} + k_{\alpha\beta} y_{\bar{x}_\beta})_{x_\alpha} + (k_{\alpha\alpha} y_{x_\alpha} + k_{\alpha\beta} y_{x_\beta})_{\bar{x}_\alpha},$$

$$\beta = 3 - \alpha, \quad \alpha = 1, 2.$$

Utilizando las fórmulas de sumación por partes (7') y (9'), obtendremos para cualesquiera $\overset{\circ}{u}$, $\overset{\circ}{v} \in \overset{\circ}{H}$

$$(\Lambda_1 \overset{\circ}{u}, \overset{\circ}{v}) = - \sum_{j=1}^{N_2-1} \sum_{i=1}^{N_1} [(k_{11} \overset{\circ}{u}_{\bar{x}_1} + k_{12} \overset{\circ}{u}_{\bar{x}_2}) \overset{\circ}{v}_{\bar{x}_1}]_{ij} h_1 h_2 - \\ - \sum_{j=1}^{N_2-1} \sum_{i=0}^{N_1-1} [(k_{11} \overset{\circ}{u}_{x_1} + k_{12} \overset{\circ}{u}_{x_2}) \overset{\circ}{v}_{x_1}]_{ij} h_1 h_2 .$$

Teniendo en cuenta, que $\overset{\circ}{v}_{\bar{x}_1}$ y $\overset{\circ}{v}_{x_1}$ son iguales a cero para $j = N_2$ y $j = 0$ respectivamente, la igualdad obtenida se puede escribir en la forma

$$(\Lambda_1 \overset{\circ}{u}, \overset{\circ}{v}) = - \sum_{j=1}^{N_1} \sum_{i=1}^{N_2} [(k_{11} \overset{\circ}{u}_{\bar{x}_1} + k_{12} \overset{\circ}{u}_{\bar{x}_2}) \overset{\circ}{v}_{\bar{x}_1}]_{ij} h_1 h_2 - \\ - \sum_{j=0}^{N_2-1} \sum_{i=0}^{N_1-1} [(k_{11} \overset{\circ}{u}_{x_1} + k_{12} \overset{\circ}{u}_{x_2}) \overset{\circ}{v}_{x_1}]_{ij} h_1 h_2 . \qquad (75)$$

Análogamente hallamos

$$(\Lambda_2 \overset{\circ}{u}, \overset{\circ}{v}) = - \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} [(k_{22} \overset{\circ}{u}_{\bar{x}_2} + k_{21} \overset{\circ}{u}_{\bar{x}_1}) \overset{\circ}{v}_{\bar{x}_2}]_{ij} h_1 h_2 - \\ - \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} [(k_{22} \overset{\circ}{u}_{x_2} + k_{21} \overset{\circ}{u}_{x_1}) \overset{\circ}{v}_{x_2}]_{ij} h_1 h_2 . \qquad (76)$$

Sumando (75) y (76), obtenemos

$$(\Lambda \overset{\circ}{u}, \overset{\circ}{v}) = -0{,}5 \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} h_1 h_2 \Big(\sum_{\alpha, \beta=1}^{2} k_{\alpha\beta} \overset{\circ}{u}_{\bar{x}_\alpha} \overset{\circ}{v}_{\bar{x}_\beta} \Big)_{ij} - \\ -0{,}5 \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} h_1 h_2 \Big(\sum_{\alpha, \beta=1}^{2} k_{\alpha\beta} \overset{\circ}{u}_{x_\alpha} \overset{\circ}{v}_{x_\beta} \Big)_{ij} . \qquad (77)$$

De aquí se deduce, que a la condición $k_{12} = k_{21}$ se cumple la igualdad

$$(\Lambda \overset{\circ}{u}, \overset{\circ}{v}) = (\overset{\circ}{u}, \Lambda \overset{\circ}{v}).$$

En virtud de la igualdad $(Au, v) = -(\Lambda \overset{\circ}{u}, \overset{\circ}{v})$ *el operador A es autoconjugado en H.*

Hallemos las cotas del operador A. Sustituyamos en (77) la función reticular $\overset{\circ}{u}$ en lugar de $\overset{\circ}{v}$ y tengamos en cuenta la condición de elipticidad (73) y la condición $\overset{\circ}{u}(x) = 0$

para $x \in \gamma$. Obtendremos

$$-(\Lambda\overset{\circ}{u}, \overset{\circ}{u}) \geqslant 0{,}5c_1 \{\sum_{j=1}^{N_2-1} h_2 [\sum_{i=1}^{N_1} (\overset{\circ}{u}_{\bar{x}_1})^2_{ij} h_1 + \sum_{i=0}^{N_1-1} (\overset{\circ}{u}_{x_1})^2_{ij} h_1] +$$

$$+ \sum_{i=1}^{N_1-1} h_1 [\sum_{j=1}^{N_2} (\overset{\circ}{u}_{\bar{x}_2})^2_{ij} h_2 + \sum_{j=0}^{N_2-1} (\overset{\circ}{u}_{x_2})^2_{ij} h_2]\} =$$

$$= c_1 [\sum_{j=1}^{N_2-1} \sum_{i=1}^{N_1} (\overset{\circ}{u}_{\bar{x}_1})^2_{ij} h_1 h_2 + \sum_{i=1}^{N_1-1} \sum_{j=1}^{N_2} (\overset{\circ}{u}_{\bar{x}_2})^2_{ij} h_1 h_2] =$$

$$= c_1(-\mathscr{R}\overset{\circ}{u}, \overset{\circ}{u}),$$

donde $\mathscr{R}$ es el operador de Laplace de diferencias. Análogamente hallaremos

$$-(\Lambda\overset{\circ}{u}, \overset{\circ}{u}) \leqslant c_2 (-\mathscr{R}\overset{\circ}{u}, \overset{\circ}{u}).$$

Teniendo en cuenta la estimación (70), obtenemos las siguientes desigualdades para el operador A:

$$\begin{aligned} c_1 (Ru, u) \leqslant (Au, u) \leqslant c_2 (Ru, u), \\ c_1\delta (u, u) \leqslant (Au, u) \leqslant c_2\Delta (u, u), \end{aligned} \tag{78}$$

donde δ y Δ están definidos en (64). Por consiguiente, el operador A correspondiente al operador elíptico de diferencias con derivadas mixtas, y el operador R, correspondiente al operador de Laplace de diferencias son energéticamente equivalentes con constantes c_1 y c_2, que no dependen del número de nodos de la red. El operador A tiene las cotas $c_1\delta = O(1)$ y $c_2\Delta = O(1/h^2)$, $(h^2 = h_1^2 + h_2^2)$, y si el número de nodos de la red es grande, entonces el operador A está mal condicionado.

Señalemos que las desigualdades (78) se mantienen ciertas aún en el caso, cuando para la aproximación de un operador diferencial L se utilizan los operadores de diferencias:

$$\Lambda y = \frac{1}{2} \sum_{\alpha=1}^{2} [(k_{\alpha\alpha} y_{\bar{x}_\alpha})_{x_\alpha} + (k_{\alpha\alpha} y_{x_\alpha})_{\bar{x}_\alpha}] +$$

$$+ \frac{1}{2} \sum_{\alpha \neq \beta}^{1 \div 2} [(k_{\alpha\beta} y_{x_\beta})_{x_\alpha} + (k_{\alpha\beta} y_{\bar{x}_\beta})_{\bar{x}_\alpha}]$$

ó

$$\Lambda y = \frac{1}{2} \sum_{\alpha=1}^{2} [(k_{\alpha\alpha} y_{\bar{x}_\alpha})_{x_\alpha} + (k_{\alpha\alpha} y_{x_\alpha})_{\bar{x}_\alpha}] +$$

$$+ \frac{1}{4} \sum_{\alpha \neq \beta}^{1 \div 2} [(k_{\alpha\beta} y_{\bar{x}_\beta})_{x_\alpha} + (k_{\alpha\beta} y_{x_\beta})_{\bar{x}_\alpha} + (k_{\alpha\beta} y_{x_\beta})_{x_\alpha} + (k_{\alpha\beta} y_{\bar{x}_\beta})_{\bar{x}_\alpha}].$$

§ 3. Conceptos fundamentales de la teoría de los métodos iterativos

1. Método de establecimiento. Más arriba se mostró que los esquemas de diferencias para ecuaciones elípticas se escriben de manera natural en forma de una ecuación operacional de primer género

$$Au = f \tag{1}$$

del operador A, que actúa en un espacio de Hilbert H de dimensión finita. A las ecuaciones lineales elípticas corresponden los operadores A lineales, mientras que a las cuasilineales corresponden los A no lineales.

La teoría de los métodos iterativos para la ecuación operacional (1) puede ser expuesta como una de las partes de la teoría general de estabilidad de los esquemas de diferencias. Los esquemas iterativos se pueden interpretar como métodos de establecimiento para la respectiva ecuación no estacionaria. Expliquemos esto en el ejemplo de una ecuación con un operador A autoconjugado, definido positivo y acotado, $A = A^* \geqslant \delta E$, $\delta > 0$.

Sea $v = v(t)$ la función abstracta de t con valores en H, es decir, para cada t fijo $v(t)$ es un elemento del espacio H. Examinemos el problema de Cauchy abstracto:

$$\frac{dv}{dt} + Av = f, \quad t > 0, \quad v(0) = v_0 \in H. \tag{2}$$

Mostremos que, $\lim\limits_{t \to \infty} \| v(t) - u \| = 0$, donde u es la solución de la ecuación (1), es decir, al crecer t la solución $v(t)$ de la ecuación no estacionaria (2) tiende a la solución u de la ecuación estacionaria (1), (no dependiente de t). Se dice que tiene lugar un «establecimiento» o «salida a un régimen estacionario». Para el error $z(t) = v(t) - u$ tene-

mos la ecuación homogénea:

$$\frac{dz}{dt}+Az=0, \quad t>0, \quad z(0)=v(0)-u.$$

Multiplicando esta ecuación escalarmente por z: $\left(\frac{dz}{dt}, z\right)+(Az, z)=0$ y teniendo en cuenta, que

$$\left(\frac{dz}{dt}, z\right)=\frac{1}{2}\frac{d}{dt}(z, z)=\frac{1}{2}\frac{d}{dt}\|z\|^2,$$

$$(Az, z)\geqslant\delta\|z\|^2,$$

obtenemos

$$\frac{d}{dt}\|z(t)\|^2+2\delta\|z(t)\|^2\leqslant 0.$$

Después de multiplicar esta desigualdad por $e^{2\delta t}>0$ tenemos

$$\frac{d}{dt}e^{2\delta t}\|z(t)\|^2\leqslant 0,$$

de donde se deduce que $e^{2\delta t}\|z(t)\|^2\leqslant\|z(0)\|^2$ ó

$$\|v(t)-u\|\leqslant e^{-\delta t}\|v(0)-u\|\to 0 \text{ siendo } t\to\infty.$$

De esta forma, resolviendo la ecuación (2) con todo $v_0\in H$, nosotros obtendremos para t suficientemente grande una solución aproximada de la ecuación (1) con cualquier exactitud prefijada. Este método para obtener la solución se llama *método de establecimiento*. Una propiedad análoga de atenuación de los datos iniciales la poseen los análogos de diferencias de la ecuación (2).

2. **Esquemas iterativos.** Detengámonos primeramente en la característica general del concepto de esquema iterativo. Supongamos que se exige hallar la solución de la ecuación (1). Al principio supondremos que A es un operador lineal definido en H.

En todo método iterativo de resolución de la ecuación (1) se parte de una cierta aproximación inicial $y_0\in H$ y sucesivamente se determinan las soluciones aproximadas $y_1, y_2, \ldots, y_k, y_{k+1}, \ldots$, donde k es el número de la iteración. La aproximación y_{k+1} se expresa mediante las aproximaciones anteriores conocidas mediante la fórmula recurrente

$$y_{k+1}=F_k(y_0, y_1, \ldots, y_k),$$

donde F_k es una cierta función dependiente, en general, del operador A del segundo miembro de f y del número de la iteración k.

Se dice que el método iterativo tiene orden m, si cada subsiguiente aproximación depende solamente de las m anteriores, es decir,

$$y_{k+1} = F_k(y_{k-m+1}, y_{k-m+2}, \ldots, y_k).$$

Los esquemas iterativos de orden alto exigen para su realización de memorización de un gran volúmen de información intermedia y por eso en la práctica con frecuencia se limitan a los valores $m = 1$ ó $m = 2$.

De la elección de la función F_k depende la estructura del esquema iterativo. Si la función es lineal, entonces el método iterativo también se llama lineal. Si F_k no depende del número de la iteración k, entonces el método iterativo se llama estacionario.

Examinemos la forma general de un esquema iterativo lineal de primer orden. Cualquier dicho esquema, en correspondencia con la definición, puede ser escrito de siguiente modo:

$$y_{k+1} = S_{k+1}y_k + \tau_{k+1}\varphi_{k+1}, \quad k = 0, 1, \ldots, \qquad (3)$$

donde los S_k son los operadores lineales, definidos sobre H y los τ_k son algunos parámetros numéricos.

En general a los esquemas iterativos se le plantea una exigencia natural: para cualquier f la solución $u = A^{-1}f \in H$ de la ecuación (1) debe ser un punto fijo del proceso de aproximaciones sucesivas (3), es decir

$$A^{-1}f = S_{k+1}A^{-1}f + \tau_{k+1}\varphi_{k+1}. \qquad (4)$$

De aquí se desprende que si ponemos

$$S_{k+1} = E - \tau_{k+1}B_{k+1}^{-1}A, \quad \varphi_{k+1} = B_{k+1}^{-1}f, \qquad (5)$$

donde B_{k+1} es un operador lineal inversible que actúa en H, entonces la condición (4) será cumplida. Sustituyendo (5) en (3), obtendremos como resultado de transformaciones no complejas

$$B_{k+1}\frac{y_{k+1}-y_k}{\tau_{k+1}} + Ay_k = f, \quad k = 0, 1, \ldots, \quad y_0 \in H. \qquad (6)$$

Conservando la terminología de la teoría de los esquemas de diferencias (véase A. A. Samarski, Teoría de esquemas de diferencias, 1977, cap. V), llamaremos a (6) forma canónica de un esquema iterativo de dos capas. Así, todo proceso iterativo lineal de primer orden puede ser escrito en la forma (6). Si $B_{k+1} \equiv E$, entonces el *esquema* iterativo *se llama*

explícito, puesto que en este caso la aproximación y_{k+1} se encuentra con ayuda de la fórmula explícita:

$$y_{k+1} = y_k - \tau_{k+1}(Ay_k - f), \quad k = 0, 1, \ldots$$

Si al menos para un k el operador B_k es diferente del operador unidad, entonces *el esquema se llama implícito*. Los números τ_k se llaman parámetros iterativos. Si τ_{k+1} depende de la aproximación iterativa y_k, entonces *el proceso iterativo será no lineal*. Es evidente, que en un proceso iterativo estacionario los operadores B_k y los parámetros τ_k (más exactamente, B_k/τ_{k+1}) no deben depender del número k de la iteración.

Señalemos, que el esquema (6) puede interpretarse como un esquema implícito de dos capas para la ecuación no estacionaria:

$$B(t)\frac{dv}{dt} + Av = f, \quad t > 0, \quad v(0) = y_0,$$

más general que la ecuación (2) examinada más arriba. Con esto el parámetro τ_{k+1} puede ser considerado como un paso por el tiempo ficticio.

La diferencia entre los esquemas iterativos y los esquemas para los problemas no estacionarios del tipo (2) consiste en lo siguiente:

1) para cualesquiera B_{k+1} y τ_{k+1} la solución u de la ecuación inicial (1) satisface (6);

2) la elección de los parámetros τ_{k+1} y de los operadores B_{k+1} se debe subordinar únicamente a las exigencias de convergencia de las iteraciones y del mínimo de operaciones aritméticas, necesarias para encontrar la solución de la ecuación (1) con una exactitud dada (para los problemas no estacionarios la elección del paso está subordinada, ante todo a la exigencia de la aproximación).

Más arriba se supuso que el operador A es lineal. Es evidente que el esquema (6) puede ser utilizado para encontrar la solución aproximada de la ecuación (1) aún en el caso del operador A no lineal. En este caso, frecuentemente, se elige el operador B_{k+1} lineal.

Los esquemas iterativos de dos capas (6) son los más usados. Sin embargo al resolver la ecuación (1) se utilizan también esquemas de tres capas, que describen los procesos iterativos de segundo orden. Los más investigados son los esquemas de tres capas de tipo «estándar». Ellos se escriben

en la forma

$$B_{k+1}y_{k+1} = \alpha_{k+1}(B_{k+1} - \tau_{k+1}A)\,y_k +$$
$$+ (1 - \alpha_{k+1})\,B_{k+1}y_{k-1} + \alpha_{k+1}\tau_{k+1}f \qquad (7)$$

para $k = 1, 2, \ldots$. Aquí se utilizan dos sucesiones de parámetros iterativos $\{\tau_k\}$ y $\{\alpha_k\}$. Para la realización del esquema (7) es necesario, además de la aproximación inicial y_0, prefijar otra aproximación y_1. Por lo común, ella se encuentra mediante y_0 empleando el esquema de dos capas (6), es decir

$$B_1y_1 = (B_1 - \tau_1A)\,y_0 + \tau_1f, \quad y_0 \in H. \qquad (8)$$

Se puede mostrar, que para (7), (8) la solución u de la ecuación (1) es un punto fijo.

Si $B_k \equiv E$ para todos los $k = 1, 2, \ldots$, entonces el esquema (7) se llama *explícito*:

$$y_{k+1} = \alpha_{k+1}(E - \tau_{k+1}A)\,y_k + (1 - \alpha_{k+1})\,y_{k-1} +$$
$$+ \alpha_{k+1}\tau_{k+1}f.$$

En el caso contrario el esquema (7) es *implícito*.

3. **Convergencia y número de iteraciones.** La diferencia fundamental entre los métodos iterativos y los directos consiste en que los primeros dan la solución exacta de la ecuación (1) solamente como el límite de la sucesión de aproximaciones iterativas $\{y_k\}$ para $k \to \infty$. Una excepción la componen los métodos de iteraciones «finitas», a los cuales pertenecen los métodos de direcciones conjugadas, que permiten teóricamente hallar la solución exacta para toda aproximación inicial, en un número finito de operaciones, si A es un operador lineal en un espacio de dimensión finita.

Para caracterizar la desviación de la aproximación iterativa y_k respecto a la solución exacta u del problema (1) se introduce *el error* $z_k = y_k - u$. El proceso iterativo se llama *convergente en el espacio energético* H_D, si $\| z_k \|_D \to 0$ cuando $k \to \infty$. Aquí H_D es el espacio generado por un operador D autoconjugado y definido positivo en H.

El sentido de introducir el espacio energético H_D consiste en lo siguiente. Como nosotros conocemos, una sucesión de elementos de H, que converja en una norma, converge también en una norma equivalente. Por eso para investigar un esquema iterativo concreto es cómodo elegir un tal espacio energético H_D, en el cual los operadores A y B_k del esquema

iterativo posean las propiedades dadas, por ejemplo sean autoconjugados y positivos definidos.

Una de las características cuantitativas más importantes de un método iterativo es el número de iteraciones. Frecuentemente se da una cierta exactitud $\varepsilon > 0$, con la cual es necesario hallar la solución aproximada de la ecuación (1). Si $\| u \|_D = O(1)$, entonces se exige, que se cumpla la condición

$$\| y_n - u \|_D \leqslant \varepsilon \quad \text{para} \quad n \geqslant n_0 (\varepsilon). \tag{9}$$

Aquí $n_0(\varepsilon)$ es el número mínimo de iteraciones, que garantizan la exactitud prefijada ε. Este número depende de la aproximación inicial que hemos tomado. Podemos aprovecharnos de la condición (9) para determinar el momento de finalizar las iteraciones, si la norma indicada puede ser calculada efectivamente en el proceso de las iteraciones. Por ejemplo, si el operador A es no degenerado y definido positivo, entonces, eligiendo en calidad de D el operador A^*A, obtendremos de (9)

$$\| y_n - u \|_D = \| A y_n - f \| \leqslant \varepsilon,$$

ya que

$$(y_n - u, y_n - u)_D = (A^*A (y_n - u), y_n - u) =$$
$$= (Ay_n - Au, Ay_n - Au) = \| Ay_n - f \|^2.$$

En el caso general para comparar la calidad de los diferentes métodos se utiliza el número de iteraciones, determinado de la condición

$$\| y_n - u \|_D \leqslant \varepsilon \| y_0 - u \|_D \text{ para } n \geqslant n_0 (\varepsilon). \tag{10}$$

Este número indica, cuántas iteraciones es suficiente cumplir para que a cualquier aproximación inicial y_0 la norma del error inicial en H_D sea disminuído en $1/\varepsilon$ veces. La condición (10) también se puede utilizar en calidad de criterio para finalizar el proceso de iteraciones.

A la ecuación (1) se le puede poner en correspondencia un gran número de esquemas iterativos (6) ó (7), (8) con cualesquiera B_k y τ_k, α_k. Entre tanto, al resolver un problema concreto surge el problema de elegir un esquema. Desde el punto de vista de la matemática numérica lo más importante es la construcción de métodos iterativos, tales que permitan obtener la solución de (1) con una exactitud dada por un tiempo de máquina mínimo. Esta exigencia de rendimiento económico del método es natural. Al efectuar las estimacio-

nos teóricas de la calidad del método dicha exigencia se sustituye con frecuencia por la exigencia de mínimo del número de operaciones aritméticas $Q(\varepsilon)$, suficientes para obtener la solución con prefijada exactitud.

El volúmen total de cálculos $Q(\varepsilon)$ es igual a $Q(\varepsilon) = \sum_{k=1}^{n} q_k$, donde q_k es el número de operaciones para el cómputo de la iteración de número k, y n es el número de iteraciones $n \geqslant n_0(\varepsilon)$. El problema de construir el método iterativo se plantea así (para el esquema (6) de dos capas): el operador A está fijo, y es necesario elegir los parámetros $\{\tau_k, k = 1, 2, \ldots, n\}$ y los operadores B_k de la condición de mínimo de $Q(\varepsilon)$.

En tal planteamiento general es poco probable que este problema tenga solución. Con frecuencia el conjunto de operadores B_k se prefija a priori, y si el número de operaciones necesarias para invertir el operador B_k, no depende de k, entonces $q_k \equiv q$ y $Q(\varepsilon) = qn_0(\varepsilon)$. En este caso el problema sobre el mínimo de $Q(\varepsilon)$ se reduce al problema de elegir los parámetros iterativos τ_k de la condición de mínimo del número de iteraciones $n_0(\varepsilon)$.

Para establecer una jerarquía de los métodos, es necesario compararlos por algunas características. Algunas veces se utilizan estimaciones asintóticas para el número de operaciones o para el número de iteraciones cuando el número de incógnitas en el esquema de diferencias tiende al infinito. Sin embargo existe una limitación real sobre el número de incognitas al resolver las ecuaciones elípticas multidimensionales por el método de redes. Así, por ejemplo, para la ecuación tridimensional de Poisson el número medio de nodes por cada variable $N \approx 100$ nos conduce a un sistema de ecuaciones algebraicas lineales con $M = 10^6$ incógnitas. Es poco probable que resulte útil el aumento del número de nodos. Por eso ante todo es necesario comparar los métodos sobre redes reales.

4. **Clasificación de los métodos iterativos.** Los métodos iterativos se caracterizan por la estructura del esquema iterativo, por el espacio energético H_D, en el cual se investiga la convergencia del método, por el tipo de método iterativo, por la condición de terminación del proceso de iteraciones, y también por el algoritmo de realización de un paso iterativo.

Nosotros examinaremos sólo esquemas iterativos de dos

y tres capas, explícitos o implícitos, para los cuales la condición de terminación del proceso de iteraciones será

$$\| y_n - u \|_D \leqslant \varepsilon \| y_0 - u \|_D, \quad \varepsilon > 0.$$

En la teoría general de los métodos iterativos se examinan métodos de dos tipos: los que utilizan información a priori sobre los operadores del esquema iterativo y los que no la utilizan (métodos de tipo variacional). En el primer caso los parámetros iterativos τ_k para el esquema (6) y τ_k, α_k para el esquema (7), (8) se eligen de la condición de mínimo de la norma del operador de permiabilidad (operador que relaciona las aproximaciones inicial y final) o de la norma del operador de transición de una iteración a otra. Con esto los parámetros iterativos se eligen de manera tal que se garantice la más alta velocidad de convergencia para la aproximación inicial más mala. En los métodos de este tipo no se utiliza la calidad de la aproximación inicial.

En los métodos de tipo variacional los parámetros iterativos se eligen de la condición de mínimo de ciertos funcionales relacionados con la ecuación inicial. Por ejemplo, en calidad de funcional se toma la norma energética del error de la k-ésima iteración. En este caso los parámetros iterativos dependen de las aproximaciones iterativas anteriores y poseen la propiedad de ser función de la aproximación inicial.

En la teoría general de los métodos iterativos nosotros renunciaremos al estudio de la estructura concreta del esquema iterativo. La teoría utiliza el mínimo de información de carácter funcional general respecto a los operadores. Esto permite alcanzar el objetivo principal, es decir indicar los principios generales para construir métodos iterativos optimos en función del carácter y del tipo de información a priori sobre el problema, y también de aquellas exigencias que se le plantean al procedimiento de resolución de este problema. Estas exigencias complementarias pueden, por ejemplo, consistir en que es necesario construir un método óptimo no para un solo problema sino para una serie de problemas con un mismo operador A, pero con miembros derechos diferentes.

Sin duda, el tener en cuenta la estructura del operador del problema a resolver permite construir métodos iterativos especiales, los cuales posean una velocidad de convergencia más alta, que los métodos de la teoría general. Esto se alcanza con una elección especial de los operadores B_k y de los

parámetros iterativos. Los métodos especiales tienen un dominio de aplicación estrecho.

Detengámonos ahora en el papel de los operadores B_k. Para los esquemas iterativos implícitos la elección de los operadores B_k debe estar subordinada a dos exigencias: garantizar la convergencia más rápida del método y la sencillez y economía en la inversión de estos operadores. Estas exigencias son contradictorias. En efecto, si en el esquema (6) tomamos $B_1 = A$ y $\tau_1 = 1$, entonces para cualquier aproximación inicial la solución de la ecuación (1) puede ser obtenida en una iteración.

En este caso la velocidad de convergencia es máxima, sin embargo la inversión de este operador B_1 es equivalente a la resolución del problema inicial.

Resulta, y esto será mostrado más abajo, que no hay necesidad de elegir el operador B_k igual al operador A. Es suficiente que sean cercanas las energías de estos operadores. Estas exigencias abren la posibilidad de elegir entre los operadores B, cercanos por su energía al operador A, operadores fácilmente inversibles.

En la actualidad se utiliza con más frecuencia el siguiente enfoque en la construcción de métodos iterativos implícitos. El operador B_{k+1} se define constructivamente en forma explícita, o bien la aproximación iterativa y_{k+1} se encuentra como resultado de algún procedimiento numérico auxiliar, el cual se puede interpretar como una inversión implícita del operador B_{k+1}.

En el primer caso por lo común se elige el operador B_{k+1} en la forma de un producto de cierto número de operadores fácilmente inversibles de modo que en cierto sentido él esté cercano al operador A. Con esto los mismos operadores, que entran en el producto, pueden depender de parámetros, los cuales pueden ser examinados como parámetros iterativos complementarios. Por ejemplo, si $B_k = (E + \omega_k A_1) \times \times (E + \omega_k A_2)$, donde los A_α son operadores, entonces los números ω_k son parámetros. En este caso la variabilidad del operador B_k se manifiesta únicamente en la dependencia de los parámetros ω_k indicados del número k de la iteración. En esta construcción del operador B_k se asegura la uniformidad del producto numérico para encontrar la solución aproximada en cada iteración.

Detengámonos en dos algoritmos para hallar una nueva aproximación y_{k+1} en el caso, cuando el operador B_{k+1} tiene una forma factorizada. Sea $B_{k+1} = B^1_{k+1} B^2_{k+1} \ldots$

$\ldots B^p_{k+1}$ y supongamos que y_{k+1} se encuentra mediante el esquema iterativo de dos capas (6). En el primer algoritmo se resuelve la sucesión de ecuaciones

$$B^1_{k+1}v^1 = F_{k+1}, \quad B^\alpha_{k+1}v^\alpha = v^{\alpha-1}, \quad \alpha = 2, 3, \ldots, p, \qquad (11)$$

donde $F_{k+1} = B_{k+1}y_k - \tau_{k+1}(Ay_k - f)$. Se ve que $y_{k+1} = v^p$. Cada una de las ecuaciones (11) debe resolverse fácilmente. El algoritmo no exige memorización de información intermedia, ya que siendo recibida, ésta se utiliza inmediatamente. La deficiencia del algoritmo consiste en la necesidad de calcular el elemento $B_{k+1}y_k$, lo que puede resultar un procedimiento complejo.

El segundo algoritmo tiene la forma de un esquema con corrección:

$$\begin{gathered} y_{k+1} = y_k - \tau_{k+1}v^p, \\ B^1_{k+1}v^1 = Ay_k - f, \quad B^\alpha_{k+1}v^\alpha = v^{\alpha-1}, \\ \alpha = 2, 3, \ldots, p. \end{gathered} \qquad (12)$$

En este caso se exige complementariamente memorizar la aproximación iterativa anterior y_k y conservarla mientras no sea hallada la corrección v^p.

En el segundo procedimiento de construcción del método iterativo implícito se parte, por ejemplo, de un esquema para la corrección (12), mientras que la corrección v^p se encuentra como la solución aproximada de la ecuación auxiliar

$$R_{k+1}v = r_k, \quad r_k = Ay_k - f. \qquad (13)$$

Supongamos que (13) se resuelve con ayuda de algún esquema iterativo de dos capas. Entonces el error $z^m = v^m - v$ satisface la ecuación homogénea

$$z^{m+1} = S_{m+1}z^m, \quad m = 0, 1, \ldots, p-1,$$

$$z^0 = v^0 - v,$$

donde S_{m+1} es el operador de transición de la iteración m-ésima a la $m+1$-ésima. De aquí hallamos

$$z^p = v^p - v = S_pS_{p-1} \ldots S_1z^0 = T_p(v^0 - v), \quad T_p = \prod_{m=1}^{p} S_m,$$

donde T_p es el operador de resolución o resolutivo. Sustituyendo aquí $v = R^{-1}_{k+1}r_k$ y eligiendo $v^0 = 0$, obten-

dremos

$$v^p = (E - T_p)\, R_{k+1}^{-1} r_k \quad \text{ó} \quad v^p = B_{k+1}^{-1} r_k, \tag{14}$$

donde mediante B_{k+1} está designado al operador $R_{k+1}(E - T_p)^{-1}$.

Sustituyamos (14) en (12) y hallaremos, que y_{k+1} satisface el esquema de dos capas (6) con el operador indicado B_{k+1}. Si la norma del operador T_p es pequeña, entonces el operador B_{k+1} es «cercano» al operador R_{k+1}. Por lo tanto en calidad del operador R_{k+1} es natural elegir un operador cercano a A.

